ARTIFICIAL INTELLIGENCE

From AI Literacy to Digital Trust: A Human-Centred Framework for Ethical, Privacy-Aware and Equitable AI Adoption in Higher Education

P. Ramya

Department of Computer Science, Jain College (Autonomous), Bangalore, India

Published November 2026

Abstract

The rapid integration of Artificial Intelligence (AI) and Generative AI into higher education is transforming teaching, learning, assessment, research, and information services. However, effective AI adoption requires more than technical proficiency; it demands AI literacy, ethical awareness, privacy protection, transparency, accountability, and digital trust. This study proposes a human-centred framework that connects AI literacy with responsible AI adoption in higher education. Drawing on recent research concerning AI literacy, ethical AI pedagogy, algorithmic fairness, data protection, governance, and students' ethical judgment, the study examines how technological competencies and ethical reasoning can jointly influence trustworthy AI use. Particular attention is given to privacy risks, algorithmic bias, explainability, academic integrity, human oversight, and inequalities in access to AI-enabled educational resources. The proposed framework positions students, educators, institutions, and information professionals as active stakeholders in establishing responsible AI ecosystems. It further emphasizes that AI literacy should evolve from understanding AI tools toward the critical evaluation of AI-generated information, ethical decision-making, and responsible data practices. The framework can support higher education institutions in developing AI policies, curricula, training programmes, and governance mechanisms that strengthen digital trust while promoting innovation and educational equity. The study provides a foundation for future empirical research examining the relationship between AI literacy, ethical competence, privacy awareness, and trustworthy AI adoption.

Keywords: Artificial intelligence literacy; digital trust; ethical AI; data privacy

1 Introduction

Artificial intelligence has moved from being a specialized technological capability to becoming an increasingly visible component of higher education. Universities are using AI for personalized learning, learning analytics, automated feedback, student support, information retrieval, content generation, research assistance, administrative decision-making, and academic advising. Generative AI has further accelerated this transformation by making sophisticated AI capabilities accessible through natural-language interfaces. Consequently, students and educators increasingly interact with AI not merely as a technological system but as an intermediary in learning, research, communication, and decision-making.

The increasing accessibility of AI creates substantial opportunities. AI-supported systems can assist learners in obtaining explanations, generating ideas, receiving feedback, translating materials, and identifying learning resources. Educators can use AI to support lesson planning, assessment design, content development, and administrative activities. However, these benefits are accompanied by significant ethical, social, and institutional challenges. Recent systematic research has identified

privacy invasion, data leakage, algorithmic bias, opaque or "black-box" algorithms, and algorithmic errors as major technological risks in AI-enabled education (Zhu et al., 2025). The same review identifies educational concerns including student homogenization, threats to teacher autonomy, erosion of teacher–student relationships, emotional disruption, and academic misconduct.

These concerns suggest that successful AI adoption cannot be measured simply by the number of AI tools implemented or the frequency with which students and faculty use them. The quality of AI adoption depends on whether users possess sufficient knowledge and critical capabilities to understand AI systems, recognize their limitations, evaluate outputs, protect personal information, identify bias, and make responsible decisions. This capability is increasingly discussed as AI literacy.

AI literacy has developed into a multidimensional construct. Rather than being limited to technical skills, contemporary approaches include understanding AI, using AI appropriately, evaluating AI outputs, creating with AI, and understanding ethical and societal implications (Ng et al., 2021). Recent research has further emphasized AI ethics knowledge, attitudes, competencies, and the ability to make context-sensitive judgments about AI use (Yang et al., 2025).

However, AI literacy alone may not automatically generate trust. Users may possess operational knowledge of AI tools while remaining unaware of privacy risks or algorithmic bias. Conversely, institutions may establish AI policies without ensuring that students and faculty understand how those policies should influence everyday decisions. This creates a gap between AI literacy and digital trust.

Digital trust in higher education can be understood as confidence that digital and AI-enabled systems operate in ways that are sufficiently transparent, secure, fair, accountable, privacy-aware, and aligned with human interests. Trust should not mean unquestioning acceptance of AI. Instead, trustworthy AI adoption requires informed users who can determine when an AI system should be used, when its output should be questioned, and when human judgment must prevail.

The central argument of this paper is therefore that higher education requires a transition from AI literacy to digital trust. This transition should be supported by a human-centred framework integrating AI knowledge, ethical reasoning, privacy protection, human agency, accountability, and digital equity.

2 Background and rationale

The expansion of AI in higher education has created a complex socio-technical environment. AI systems do not operate independently of people and institutions. Their outcomes are influenced by training data, algorithms, developers, institutional policies, users, infrastructure, and social contexts.

Zhu et al. (2025), in a systematic review of 75 studies, categorized ethical risks of AI in education into technological, educational, and societal dimensions. The technological dimension included privacy invasion, data leakage, algorithmic bias, black-box algorithms, and algorithmic error. The educational dimension included concerns about student development, teaching practices, teacher–student relationships, emotional disruption, and academic misconduct. At the societal level, the review highlighted digital divides, accountability gaps, and conflicts of interest.

These findings demonstrate that AI-related risks cannot be addressed through technical training alone. A student may know how to generate text using an AI tool but may not know whether entering personal information into the tool is appropriate. An educator may understand how to generate

automated feedback but may not recognize that an algorithm can reproduce biases embedded in its training data. An administrator may use predictive analytics without fully understanding the implications of automated classification.

Recent research on AI literacy education similarly demonstrates that AI literacy is an emerging and multidimensional field. Yang et al. (2025) reviewed a decade of AI literacy education research and highlighted the growing importance of structured AI literacy education. Hackl et al. (2026) proposed an "AI literacy heptagon" as a structured approach to AI literacy in higher education, reflecting the movement toward multidimensional rather than purely technical understandings of AI literacy.

The importance of ethical judgment is also becoming increasingly visible. Research involving undergraduate students has examined whether AI literacy is associated with the ability to make appropriate judgments about generative AI scenarios. Such research distinguishes ethical judgment from simple awareness: knowing that privacy or fairness matters is different from being able to apply those principles when confronted with a real situation (Reed et al., 2026).

The Indian higher-education context provides an especially relevant setting for examining these issues. Krishnan (2026) reported a three-year mixed-methods study involving 1,168 students across 12 Indian higher education institutions. The study integrated AI literacy, socio-technical ethics, and equity-oriented access. Its findings indicate that ethical AI pedagogy can be associated with substantial gains in AI literacy and ethical awareness, while infrastructure differences continue to influence outcomes. Such evidence reinforces the importance of combining educational interventions with institutional and infrastructural support.

3 Problem statement

The rapid adoption of AI in higher education has created a mismatch between technological availability and human preparedness. Universities are increasingly acquiring AI-enabled platforms, while students and faculty may lack systematic preparation for evaluating their outputs, understanding their limitations, protecting personal information, and recognizing ethical consequences.

Existing AI literacy initiatives frequently emphasize tool usage, prompting, or basic conceptual knowledge. Although these competencies are important, they do not fully address the broader conditions required for trustworthy AI adoption.

Three major problems can therefore be identified.

First, AI literacy and ethical literacy are often treated separately. Users may learn how to operate AI systems without receiving equivalent preparation in fairness, privacy, transparency, accountability, and responsible decision-making.

Second, institutional trust is difficult to establish when AI systems are opaque. Students and faculty may not know what data are collected, how those data are processed, whether information is retained, or how algorithmic decisions are generated.

Third, AI adoption can reproduce or intensify existing inequalities. Students with reliable connectivity, personal devices, paid AI services, and stronger digital support may benefit more than students who lack these resources. Therefore, AI adoption without an equity perspective may widen rather than reduce educational disparities.

The problem is consequently not whether higher education should adopt AI, but how institutions can create conditions in which AI adoption remains human-centred, ethical, privacy-aware, accountable, and equitable.

4 Research gap

The literature reviewed for this paper reveals several important gaps.

4.1 Fragmentation between AI literacy and AI ethics

Many studies conceptualize AI literacy as a collection of competencies, while ethical AI research separately examines privacy, bias, transparency, and accountability. There is a need for an integrated model connecting these dimensions.

4.2 Limited emphasis on digital trust as an outcome

AI literacy is often evaluated through knowledge, skills, attitudes, or confidence. Less attention has been given to how these competencies contribute to institutional and individual digital trust.

4.3 Insufficient integration of privacy

Privacy is frequently discussed as a governance concern, but students and educators also require practical privacy literacy: understanding what information should or should not be entered into AI systems, how data may be processed, and what institutional safeguards should exist.

4.4 Human agency requires greater emphasis

Responsible AI frameworks commonly emphasize human oversight. However, higher education requires a stronger focus on human agency—the ability of students and educators to question AI outputs, reject inappropriate recommendations, and retain responsibility for consequential decisions.

4.5 Equity remains an institutional challenge

AI literacy programs can produce unequal outcomes if infrastructure, connectivity, devices, accessibility, language support, or training opportunities differ across student groups. Evidence from Indian higher education highlights the continuing importance of infrastructure and rural–urban differences (Krishnan, 2026).

The proposed framework addresses these gaps by connecting AI literacy to ethical reasoning, privacy awareness, human agency, accountability, and equity, with digital trust positioned as the resulting institutional and individual outcome.

5 Objectives of the study

The objectives of this conceptual study are:

1. To examine the emerging dimensions of AI literacy in higher education.
2. To identify the ethical, privacy, accountability, and equity challenges associated with AI adoption.
3. To examine the relationship between AI literacy and digital trust.

4. To develop a human-centred framework for ethical and privacy-aware AI adoption in higher education.
5. To propose institutional strategies for strengthening responsible and equitable AI adoption.

6 Research questions

The study addresses the following research questions:

- **RQ1:** What dimensions of AI literacy are important for responsible AI adoption in higher education?
- **RQ2:** What ethical, privacy, accountability, and equity factors influence digital trust in AI-enabled higher education?
- **RQ3:** How can AI literacy contribute to informed human agency and responsible AI use?
- **RQ4:** How can higher education institutions operationalize a human-centred framework for trustworthy AI adoption?

7 Literature review

7.1 AI literacy

AI literacy refers to the knowledge and competencies required to understand, interact with, evaluate, and critically reflect upon AI systems. Long and Magerko (2020) established an influential foundation by conceptualizing AI literacy through competencies related to understanding AI, recognizing its capabilities and limitations, and critically evaluating AI systems.

Ng et al. (2021) further conceptualized AI literacy around dimensions including knowing and understanding AI, using AI, evaluating and creating with AI, and understanding AI ethics. This multidimensional approach is particularly relevant to higher education because students are not merely consumers of AI outputs; they are increasingly creators, evaluators, researchers, and decision-makers.

Recent scholarship indicates that AI literacy education has become increasingly important as generative AI becomes embedded in educational practices. Yang et al. (2025) identified growing research attention toward structured AI literacy education. Hackl et al. (2026) similarly proposed a structured model for higher-education AI literacy, suggesting that effective literacy should extend beyond basic tool familiarity.

Therefore, AI literacy should be interpreted as a continuum of understanding → application → evaluation → ethical judgment → responsible creation (Figure 1).



Figure 1. AI literacy as a continuum from understanding to responsible creation.

This progression provides the foundation for the framework proposed in this paper.

7.2 Ethical AI adoption

Responsible AI adoption requires attention to values such as fairness, transparency, accountability, human autonomy, privacy, and non-maleficence. UNESCO (2021) emphasized human rights, human dignity, fairness, transparency, explainability, and human oversight as important principles for ethical AI.

In education, these principles become particularly significant because AI systems can influence learning opportunities, assessment, student support, academic integrity, and institutional decision-making.

Zhu et al. (2025) demonstrate that ethical risks in education are multidimensional. Algorithmic bias, privacy invasion, data leakage, and black-box decision-making may affect both individual users and institutional processes. These risks cannot be managed solely by students or teachers; institutional governance and technological safeguards are also required.

7.3 Privacy-aware AI adoption

Privacy is central to digital trust. Educational environments contain highly sensitive information, including academic records, demographic data, learning behaviours, assessment results, communications, and potentially biometric or behavioural information.

AI systems may process user inputs, interaction histories, documents, and institutional data. Consequently, students and faculty need practical knowledge about data minimization, consent, confidentiality, secure use, and information disclosure.

Zhu et al. (2025) identified privacy invasion and data leakage as prominent technological risks in AI-enabled education. Their review emphasizes that risks can emerge during data collection, storage, sharing, and use.

Privacy awareness should therefore become a core component of AI literacy rather than an independent compliance topic.

7.4 AI bias and fairness

AI systems can reproduce or amplify biases present in training data, system design, institutional practices, or historical records. In higher education, biased systems could potentially influence admissions, assessment, recommendations, student profiling, or resource allocation.

Zhu et al. (2025) identified algorithmic bias as one of the frequently reported ethical risks in the literature. Bias may arise from incomplete datasets, historical inequalities, design choices, or human assumptions.

AI literacy should therefore include the ability to ask:

- Who created the system?
- What data were used?
- Which groups may be underrepresented?
- What assumptions are embedded in the system?
- Could the output disadvantage particular groups?
- What human review mechanisms exist?

Such questioning transforms AI literacy from operational competence into critical competence.

7.5 Human agency and accountability

Human-centred AI requires that humans remain capable of exercising meaningful judgment. In higher education, AI should support rather than replace educational relationships and professional responsibility.

The literature on responsible AI governance emphasizes human oversight and accountability. Recent work on generative AI governance argues that educators should retain professional judgment and that institutions should continuously evaluate AI's effects on learning, academic integrity, and student autonomy.

Human agency is particularly important when AI outputs affect consequential decisions. An educator should not automatically accept an AI-generated assessment recommendation, and a student should not assume that an AI-generated explanation is accurate simply because it is fluent.

The framework therefore positions human verification as a mandatory bridge between AI output and consequential educational action.

7.6 Digital equity

Digital equity concerns whether individuals have meaningful and comparable opportunities to access and benefit from digital technologies.

AI adoption can generate new forms of inequality because advanced AI tools may require reliable connectivity, modern devices, paid subscriptions, digital skills, and language competencies. Students in rural or economically constrained environments may experience different levels of access.

Krishnan (2026) provides recent evidence from Indian higher education that ethical AI pedagogy can support AI literacy and ethical awareness while infrastructure limitations continue to shape equity outcomes. This indicates that educational intervention alone cannot completely address AI inequality.

Equity must therefore be incorporated into institutional AI planning from the beginning.

8 Proposed human-centred framework

This study proposes a framework consisting of six interconnected components:

8.1 Component 1: AI literacy

The foundation involves knowledge of:

- AI concepts and terminology
- Machine learning and generative AI
- AI capabilities and limitations
- Prompting and interaction
- AI output evaluation
- Hallucination and misinformation
- AI-supported research and learning

8.2 Component 2: Ethical awareness

Users should understand:

- Fairness
- Bias
- Transparency
- Accountability
- Academic integrity
- Human dignity
- Responsible use
- Social consequences

8.3 Component 3: Privacy awareness

This dimension includes:

- Data minimization
- Consent
- Confidentiality
- Personal-data protection
- Secure AI use
- Institutional data policies
- Awareness of data retention and processing

8.4 Component 4: Human agency

Human agency involves:

- Critical questioning
- Verification of AI outputs
- Independent reasoning
- Ability to reject AI recommendations
- Human review of consequential decisions
- Maintaining teacher–student relationships

8.5 Component 5: Digital equity

Institutions should address:

- Device accessibility
- Internet connectivity
- Affordable AI access
- Accessibility for students with disabilities
- Language diversity
- Rural–urban differences
- Digital skills gaps

8.6 Component 6: Digital trust

Digital trust is proposed as the principal outcome of the framework.

Trust develops when users perceive AI systems and institutional practices as transparent, secure, fair, accountable and equitable (Figure 2).

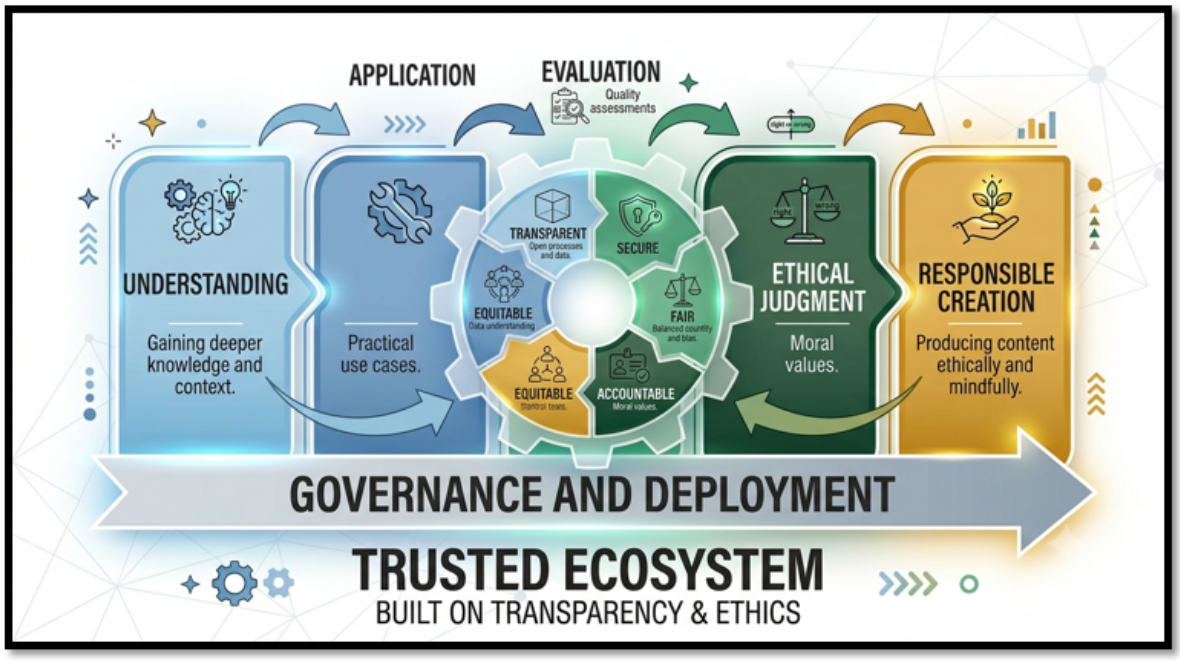


Figure 2. Attributes of trustworthy AI systems and institutional practices.

The conceptual relationship is represented in Figure 3.

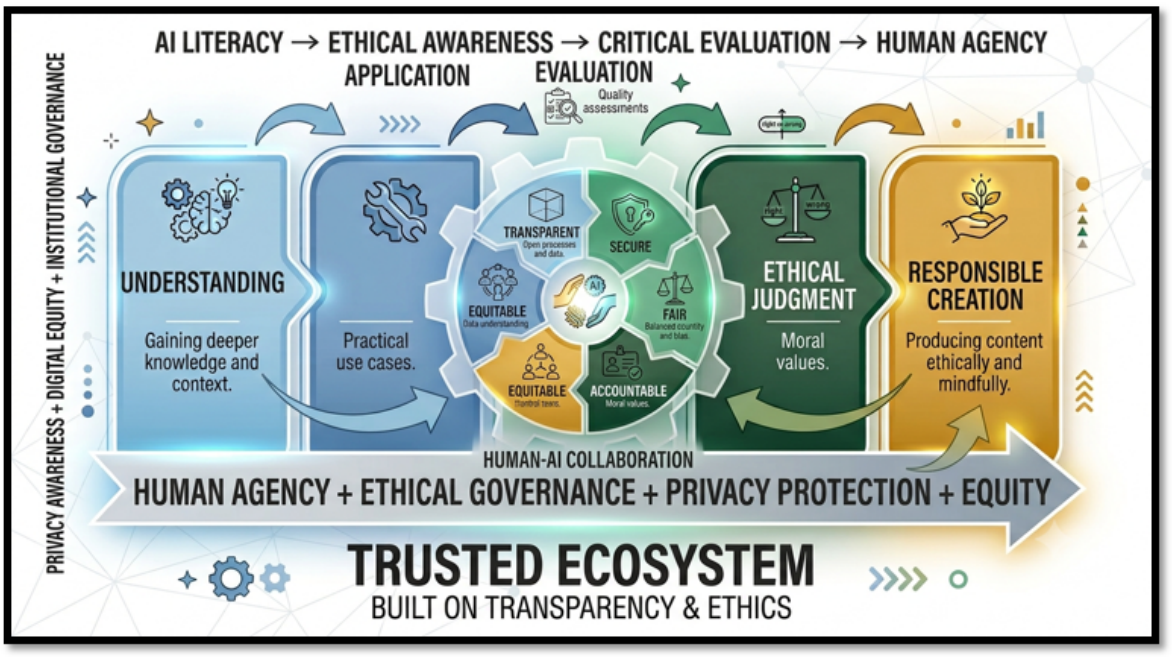


Figure 3. Conceptual relationship from AI literacy to digital trust.

Digital trust then supports responsible and sustainable AI adoption.

9 Methodology

This paper adopts a conceptual integrative literature-review methodology. The purpose is not to generate primary empirical results but to synthesize recent research and develop a theoretically grounded framework.

The literature base includes recent peer-reviewed studies addressing AI literacy, ethical AI, responsible AI governance, AI ethics education, generative AI, privacy, equity, and higher education. Particular attention was given to studies published during 2025 and 2026 because of the rapid development of generative AI and emerging empirical evidence.

The analytical process consisted of four stages.

9.1 Stage 1: Identification

Research relevant to AI literacy, ethical AI adoption, responsible AI, digital trust, privacy, and equity was identified.

9.2 Stage 2: Thematic analysis

The literature was examined for recurring concepts. Six major themes emerged:

- AI literacy
- Ethical awareness
- Privacy
- Human agency

- Digital equity
- Digital trust

9.3 Stage 3: Conceptual integration

The themes were examined in relation to each other rather than as independent variables. This produced a human-centred pathway from AI literacy toward trustworthy AI adoption.

9.4 Stage 4: Framework development

The final framework integrates individual competencies with institutional governance and infrastructural conditions.

10 Proposed propositions

The framework provides the basis for future empirical testing through the following propositions (Figure 4).



Figure 4. Propositions (P1–P7) of the proposed digital trust framework.

- **P1:** Higher levels of AI literacy are positively associated with the ability to critically evaluate AI outputs.
- **P2:** AI literacy combined with ethical awareness is more strongly associated with responsible AI use than technical AI knowledge alone.
- **P3:** Privacy awareness strengthens users' confidence in institutional AI practices.
- **P4:** Human agency mediates the relationship between AI literacy and digital trust.
- **P5:** Transparent institutional AI governance positively contributes to digital trust.
- **P6:** Digital equity moderates the relationship between AI literacy interventions and responsible AI adoption.

- **P7:** Digital trust is positively associated with sustained and responsible AI adoption in higher education.

11 Discussion

The proposed framework makes an important conceptual distinction between AI use and AI readiness. A student may use ChatGPT or another generative AI system frequently without possessing the competencies necessary to evaluate its outputs. Similarly, an educator may use AI for content generation while lacking awareness of data privacy or algorithmic bias.

Therefore, frequency of AI use should not be treated as a proxy for AI literacy.

The framework also suggests that digital trust should not be understood as unconditional confidence in technology. Instead, trustworthy AI adoption requires calibrated trust. Users should understand when AI can be useful and when it requires verification or human intervention.

This is particularly important because AI systems can generate convincing but inaccurate information. Ethical AI literacy should therefore include epistemic vigilance—the capacity to distinguish plausible output from verified knowledge.

The framework also extends the concept of digital trust from an individual-level phenomenon to an institutional one. Universities influence trust through policies, procurement decisions, data practices, transparency mechanisms, staff training, accessibility measures, and grievance procedures.

For example, a university may introduce an AI-powered student-support system. Trust in the system will depend not only on whether students know how to use it but also on whether they understand what data are collected, whether they can challenge an incorrect recommendation, whether a human can intervene, and whether disadvantaged students receive comparable access.

The framework consequently adopts a socio-technical perspective. Technology, people, institutions, and infrastructure are interconnected.

Recent responsible-AI reviews similarly emphasize that governance should include transparency, fairness, privacy protection, accountability, and human control. Existing frameworks, however, are often normative and require greater empirical validation. The present framework contributes by identifying AI literacy as an important mechanism through which governance principles can become meaningful in everyday educational practice.

The Indian context adds another dimension. Krishnan's (2026) longitudinal study demonstrates that ethical AI pedagogy can be integrated with AI literacy and equity-oriented educational design. At the same time, differences in infrastructure can affect outcomes. This supports the proposition that digital equity should not be treated as an optional component of AI adoption.

Academic libraries may also have an important role. Libraries traditionally support information literacy, source evaluation, research skills, and ethical information use. These functions can be expanded to include critical AI literacy, AI-assisted research evaluation, source verification, privacy awareness, and responsible use of generative AI.

12 Implications

12.1 Implications for universities

Universities should establish institution-wide AI literacy programmes rather than relying exclusively on individual faculty initiatives. AI literacy should be integrated into orientation programmes, curricula, faculty development, research training, and professional development.

12.2 Implications for educators

Faculty members should teach students how to question AI outputs rather than simply how to generate them. Assignments can incorporate AI critique, source verification, reflection on AI use, and documentation of AI-assisted work.

12.3 Implications for students

Students should develop competencies in:

- Evaluating AI-generated information
- Recognizing hallucinations
- Protecting personal information
- Identifying bias
- Citing or disclosing AI assistance appropriately
- Maintaining independent reasoning
- Using AI as a support rather than an unquestioned authority

12.4 Implications for academic libraries

Academic libraries can develop AI literacy workshops, curate reliable AI resources, teach verification practices, support ethical research, and assist students and faculty in understanding data and privacy implications.

12.5 Implications for policymakers

Policy should address AI literacy, privacy, transparency, accountability, accessibility, and equity simultaneously. Regulations focusing exclusively on technological compliance may not adequately address user behaviour and institutional culture.

12.6 Implications for AI developers

AI developers serving higher education should prioritize transparency, explainability, privacy controls, accessibility, bias monitoring, and mechanisms that facilitate human oversight.

13 Recommendations

Based on the proposed framework, higher education institutions should consider the following measures:

1. Develop institutional AI literacy frameworks covering technical, ethical, privacy, and critical-evaluation competencies.
2. Introduce privacy-aware AI training for students, faculty, researchers, and administrators.

3. Establish AI governance committees involving academic, technical, legal, library, student, and administrative stakeholders.
4. Implement human-in-the-loop mechanisms for high-impact educational decisions.
5. Create transparent AI-use policies specifying acceptable, restricted, and prohibited uses.
6. Provide equitable AI access through institutional tools, library facilities, low-bandwidth alternatives, and accessibility support.
7. Integrate AI ethics into curricula rather than treating it solely as a separate workshop.
8. Develop AI-output verification practices for academic research and assessment.
9. Conduct regular algorithmic and privacy audits of institutional AI systems.
10. Establish grievance and appeal mechanisms through which students and faculty can challenge AI-supported decisions.
11. Measure AI literacy longitudinally rather than relying on one-time training assessments.
12. Evaluate digital trust empirically through validated measures of transparency, security, fairness, accountability, and perceived human control.

14 Limitations

This paper is conceptual and does not report primary survey or experimental data. The proposed relationships and propositions therefore require empirical validation.

Second, AI technology changes rapidly. AI literacy competencies relevant to one generation of systems may require modification as new technologies emerge.

Third, digital trust can vary according to discipline, institution, culture, technological infrastructure, and prior experience with AI. Therefore, the framework should be adapted to different higher-education contexts.

Finally, although recent literature from India provides important evidence regarding equity and ethical AI pedagogy, further research involving diverse regions, institutions, languages, disciplines, and socioeconomic groups is necessary.

15 Future research directions

Future studies should empirically test the proposed framework using quantitative and mixed-methods approaches.

Structural equation modelling could be used to test relationships between AI literacy, ethical awareness, privacy awareness, human agency, institutional governance, digital equity, and digital trust.

Longitudinal research could examine whether AI literacy interventions produce sustained changes in responsible AI behaviour.

Comparative studies could examine differences across:

- Technical and non-technical disciplines
- Undergraduate and postgraduate students
- Urban and rural institutions
- Public and private universities

- Different socioeconomic groups
- Different language backgrounds

Future research should also investigate whether digital trust predicts responsible AI adoption, continued use, or willingness to challenge AI-generated decisions.

Experimental research could compare traditional AI-skills training with integrated ethical AI literacy programmes to determine whether the latter produces stronger critical-evaluation and ethical-judgment outcomes.

16 Conclusion

Artificial intelligence presents higher education with an opportunity to transform teaching, learning, research, and administration. However, technological capability alone does not guarantee responsible adoption. The central challenge is to ensure that individuals and institutions possess the knowledge, ethical awareness, privacy understanding, critical judgment, and infrastructure required to use AI responsibly.

This paper proposed a human-centred framework that connects AI literacy to digital trust through six dimensions: AI literacy, ethical awareness, privacy awareness, human agency, digital equity, and institutional governance. The framework recognizes that trustworthy AI adoption is neither exclusively a technical issue nor an individual responsibility. It is a socio-technical process involving learners, educators, administrators, libraries, policymakers, developers, and institutional systems.

The transition from AI literacy to digital trust requires a shift from asking "Can users operate AI?" to asking "Can users understand, question, evaluate, govern, and responsibly use AI?"

The proposed framework positions digital trust not as unquestioning confidence but as informed, evidence-based and human-centred confidence accompanied by the capacity to challenge AI when necessary.

For higher education, the ultimate objective should therefore not be maximum AI adoption. It should be responsible AI adoption that strengthens learning, preserves human agency, protects privacy, promotes fairness, and expands rather than restricts educational opportunity.

References

[1] Alfiras M, Emran AQ, Mohamed AM (2025). Ethics and governance of generative AI in education: A systematic review on responsible adoption. *Discover Education*, 5(1). https://doi.org/10.1007/s44217-025-01051-y

[2] Hackl V, Müller AE, Sailer M (2026). The AI literacy heptagon: A structured approach to AI literacy in higher education. *Computers and Education: Artificial Intelligence*, 10, 100540. https://doi.org/10.1016/j.caeai.2026.100540

[3] Krishnan S (2026). Longitudinal effects of ethical AI pedagogy on AI literacy, moral competence, and educational equity in Indian higher education: A three-year mixed-methods study with implications for academic libraries (NEP 2020 alignment). *Information Services & Use*, 46(3), 76–93. https://doi.org/10.1177/18758789261452185

[4] Liu, et al. (2026). AI literacy interventions in education: A meta-analysis of effects and moderators. *Educational Research Review*.

[5] Long D, Magerko B (2020). What is AI literacy? Competencies and design considerations. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Article 191, 1–16. Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

[6] Ng DTK, Leung JKL, Chu SKW, Qiao MS (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

[7] Pinski M, Benlian A (2024). AI literacy for users – A comprehensive review and future research directions of learning methods, components, and effects. *Computers in Human Behavior: Artificial Humans*, 2(1), 100062. https://doi.org/10.1016/j.chbah.2024.100062

[8] Reed JM, Ferdig RE, Dodson TM, Gunstad J, Hughes JW (2026). AI literacy and scenario-based ethical judgment about generative AI among undergraduate students. *Journal of Academic Ethics*, 24(3). https://doi.org/10.1007/s10805-026-09757-w

[9] UNESCO (2021). Recommendation on the ethics of artificial intelligence. *UNESCO*, Paris.

[10] Yang Y, Zhang Y, Sun D, He W, Wei Y (2025). Navigating the landscape of AI literacy education: Insights from a decade of research (2014–2024). *Humanities and Social Sciences Communications*, 12. https://doi.org/10.1057/s41599-025-04583-8

[11] Zhu H, Sun Y, Yang J (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications*, 12, 1111. https://doi.org/10.1057/s41599-025-05252-6