



REVIEW ARTICLE

A Governance Framework for Operationalising Responsible AI Use in Students' Research Projects: Illustrations from Viet Nam

Dao Mai Thi Tran

Pham Van Dong University, Quang Ngai Province, Viet Nam

Email: tmdao@pdu.edu.vn

Article history

Received: 12 June, 2026

Accepted: 29 July, 2026

Published: 15 October, 2026

Keywords

Generative AI, academic integrity, responsible AI, student research, AI governance, Viet Nam

ABSTRACT

Generative artificial intelligence (AI) is increasingly used in students' research projects for literature searching, idea development, drafting, language editing, information processing, and research planning. These uses can support research learning, but they also raise academic integrity concerns, including unverifiable sources, fabricated references, undeclared AI assistance, over-reliance on AI-generated outputs, unclear authorship, and unfair assessment. This conceptual paper argues that responsible AI use in students' research projects cannot be secured through individual AI literacy or general guidance alone. Higher education institutions also need governance mechanisms that make AI use transparent, verifiable, assessable, and accountable. Using targeted literature identification and structured conceptual synthesis, the paper maps recurring governance problems to normative principles, institutional requirements, and operational mechanisms. It proposes a five-component framework corresponding to five governance functions: policy (boundary setting), transparency (traceability), assessment (evaluability), audit (assurance), and guidance/support (capability building). Viet Nam is used as an illustrative emerging AI governance context, where national policy directions on digital competence and AI are developing but operational mechanisms for student research remain under-specified. The framework and associated tools are presented as adaptable design artefacts for future institutional testing rather than as validated interventions.

1. INTRODUCTION

In this paper, students' research projects refer to research activities conducted by students within higher education institutions, including coursework research, undergraduate theses, capstone projects, and other supervised research assignments. In these projects, generative AI is increasingly used for literature searching, source summarisation, idea development, outlining, drafting, academic language editing, citation formatting, and methodological planning. These uses may support research learning, but they also raise concerns about academic integrity, authorship, source validity, student responsibility, and fairness in assessment. The central issue is therefore not simply whether students use AI, but whether such use is transparent, verifiable, traceable, and accountable.

Recent discussions of postplagiarism suggest that the AI era requires higher education to move beyond traditional understandings of plagiarism and reconsider authorship, human contribution, transparency in knowledge production, and accountability (Eaton, 2023). Academic integrity is not merely the avoidance of cheating or plagiarism; it concerns the values, norms, and behaviours that shape learning, teaching, research, assessment, and professional responsibility (Macfarlane et al., 2014). A paper may contain no copied text yet still raise integrity concerns.

Problems arise when students use AI-generated sources without verification, include fabricated references, adopt AI-generated reasoning without critique, or conceal the extent of AI assistance.

International frameworks on AI literacy and competence emphasise the knowledge, skills, and judgement that students and educators need to use AI responsibly (Long & Magerko, 2020; Miao & Cukurova, 2024; Miao et al., 2024; Ng et al., 2021). Individual capability, however, does not establish shared institutional rules, documentation requirements, assessment criteria, or accountability mechanisms. The Artificial Intelligence Assessment Scale (AIAS) clarifies the degree and pedagogical role of AI use within assessment (Perkins et al., 2024). Policy analyses also show growing institutional attention to academic integrity, stakeholder responsibilities, implementation, and assessment practice (Chan, 2023; Luo, 2024; An et al., 2025). These issues are particularly relevant to student research, where competence is demonstrated through problem formulation, literature review, argumentation, method selection, data handling, interpretation, and citation practice.

The contribution of the proposed framework lies in the institutional linkage among mechanisms, not in any mechanism considered alone. AI literacy frameworks describe individual capabilities; AIAS structures the pedagogical role of AI in assessment; and broader governance approaches address organisation-wide principles, risks, and implementation. This paper addresses the narrower problem of governing generative AI across the student research process. It links boundaries, process evidence, evaluation, selective assurance, and support within a domain-specific governance architecture.

In Viet Nam, recent policy developments show increasing attention to digital competence, digital transformation, and AI. The National Strategy for Artificial Intelligence Research, Development and Application through 2030 was issued under Decision No. 127/QĐ-TTg (Prime Minister of Viet Nam, 2021). The Digital Competency Framework for Learners and the Digital Literacy and Skills Development Programme for Students in Higher Education Institutions provide a policy basis for developing learners' digital capacities (Ministry of Education and Training, 2025a, 2025b). The Viet Nam National Forum on Artificial Intelligence in Education also indicates that AI has become a priority area for educational capacity building (UNICEF Viet Nam, 2025). Yet these policies provide broad direction rather than operational mechanisms for research projects conducted by students.

Viet Nam is used in this paper as an illustrative emerging AI governance context rather than as an empirical national case study. The relevance of this context lies in the gap between broad policy directions on digital competence, AI, and digital transformation and the more specific mechanisms needed to guide AI use in students' research projects. This gap is not unique to Viet Nam. The framework is not intended to replace national policy or disciplinary judgement; rather, it is positioned primarily at the institutional level, where broad principles must be translated into procedures for supervision, documentation, assessment, and review.

Recent studies in Viet Nam also show that AI and academic integrity are becoming salient issues. Ngo et al. (2024) documented Vietnamese EFL teachers' perceptions of academic integrity challenges in the age of generative AI. Le et al. (2025) reported both perceived benefits and threats associated with AI tools among the Vietnamese students included in their study. Do et al. (2026) developed and validated an AI literacy scale for university students in Viet Nam. These studies are important, but they mainly address perceptions, capacities, or behaviours. What remains underdeveloped is an institutional governance framework that can guide, document, assess, and audit students' use of AI in research.

This paper proposes such an architecture. It is limited to conceptual framework development and does not report empirical findings. The associated operational tools are design artefacts rather than validated instruments. The guiding question is: Which institutional functions are needed to govern students' use of generative AI throughout the research process, and how should those functions be organised?

2. CONCEPTUAL FOUNDATIONS

The framework draws on four bodies of work: academic integrity, responsible and trustworthy AI, AI literacy, and assessment scholarship.

2.1. Academic integrity

Academic integrity provides the normative foundation. In the context of generative AI, integrity must extend from the final product to the process through which that product is created. A research report may appear original

while still being problematic if AI-generated claims, references, or arguments are used without verification and critical judgement. Governance should therefore focus not only on detecting misconduct after submission, but also on designing processes that make transparency, verification, and scholarly responsibility part of research practice (Macfarlane et al., 2014; Eaton, 2023; Cotton et al., 2024).

2.2. Responsible and trustworthy AI

Responsible and trustworthy AI provides cross-cutting principles that can be translated into educational governance. These include human oversight, transparency, privacy and data governance, fairness, accountability, and the preservation of human agency (European Commission High-Level Expert Group on AI, 2019; Miao & Holmes, 2023). In students' research projects, these principles must inform decisions about acceptable use, disclosure, source verification, assessment, and proportionate review rather than remain general ethical statements.

2.3. AI literacy and institutional capability

AI literacy concerns the competences that students and educators need to engage with AI critically and responsibly. Students should understand AI capabilities and limitations, evaluate outputs, and exercise ethical judgement, while educators need sufficient competence to interpret policies and support responsible use (Long & Magerko, 2020; Miao & Cukurova, 2024; Miao et al., 2024; Ng et al., 2021). Individual capability, however, does not create shared institutional rules, evidence requirements, assessment criteria, or assurance mechanisms. Responsible AI use therefore requires institutional capability as well as individual competence.

2.4. Assessment design

Assessment design is central to AI governance. If AI can support drafting, summarising, language editing, coding, and argument generation, the final written product alone may not adequately demonstrate students' research competence. AIAS clarifies appropriate levels of AI use within assessment (Perkins et al., 2024). Analyses of university guidance and policy likewise show the need to align permitted AI use with assessment purpose and academic integrity (Moorhouse et al., 2023; Luo, 2024). Recent scholarship further argues that policy statements are insufficient when generative AI changes the capabilities that a task can validly evidence. This strengthens the case for structural assessment redesign and process-based evidence (Corbin et al., 2025).

2.5. From responsible AI use to institutional governability

The four conceptual foundations address different dimensions of responsible AI use but do not, individually, constitute a complete governance architecture for students' research projects. Academic integrity provides normative expectations; responsible and trustworthy AI provides cross-cutting principles; AI literacy defines individual capabilities; and assessment scholarship provides mechanisms for evaluating learning and competence. The conceptual gap addressed in this paper lies in connecting these perspectives to institutional procedures that govern AI use throughout the student research process.

This paper therefore distinguishes between responsible AI use and institutional governability. Responsible use concerns whether students exercise appropriate judgement, verification, transparency, and human agency when using AI. Institutional governability concerns whether higher education institutions have mechanisms through which such use can be consistently bounded, documented, interpreted, reviewed, and supported. Individual responsibility remains necessary, but responsible practice becomes institutionally governable only when expectations are translated into procedures, evidence, and shared decision rules.

The synthesis distinguishes five governance needs: common boundaries, traceable process evidence, evaluation of competence, selective assurance, and implementation support. These needs underpin the functions specified in the framework.

The functions operate as a cycle: implementation, assessment, and audit generate feedback for revising policy, guidance, and assessment criteria.

To clarify the contribution of the proposed framework, Table 1 positions it in relation to existing approaches that inform the governance of AI use in student research. The framework does not replace AI literacy, academic integrity, trustworthy AI, or assessment-oriented approaches. Rather, it integrates these perspectives into an operational governance structure for research projects conducted by students.

Table 1. Positioning the proposed framework in relation to existing approaches

Existing approach	Primary focus	Remaining gap for student research governance	Contribution of the proposed framework
Academic integrity scholarship	Norms of honesty, authorship, responsibility, fairness, and misconduct prevention	May not specify operational procedures for documenting, assessing, and reviewing AI-supported research processes	Extends integrity from the final product to the research process by linking disclosure, verification, assessment, and accountability
AI literacy and competency frameworks	Individual learner and educator capability to understand, evaluate, and use AI responsibly	Individual capability alone does not create shared institutional rules, process evidence, assessment criteria, or assurance mechanisms	Embeds individual capability within institutional procedures for policy, documentation, assessment, verification, and support
Trustworthy and responsible AI frameworks	Cross-cutting principles such as transparency, human oversight, privacy, fairness, and accountability	Principles are often too broad to specify procedures for student supervision, research submission, and assessment	Translates cross-cutting responsible-AI principles into governance functions and operational mechanisms for student research
AI assessment frameworks such as AIAS	Assessment tasks and the degree or pedagogical role of AI integration	Do not by themselves constitute a full governance cycle spanning policy, process documentation, selective verification, support, and institutional learning	Connects assessment with policy, transparency, assurance, and capability building across the student research process
General institutional AI governance and policy approaches	Organisation-wide AI use, policy, risk, responsibility, and implementation	May be too broad to specify procedures tailored to students' research projects and their supervision and assessment	Provides a domain-specific governance architecture for AI use across supervision, preparation, submission, assessment, and review

3. APPROACH TO FRAMEWORK DEVELOPMENT

3.1. Conceptual design and source identification

This paper adopts a conceptual and design-oriented approach to framework development. It does not evaluate the effectiveness of an intervention, measure changes in student competence, or report empirical survey, training, or implementation outcomes. Instead, it develops and refines a governance framework by synthesising concepts relevant to responsible generative AI use in students' research projects and translating these concepts into institutional governance functions and operational mechanisms.

The approach is best understood as targeted literature identification combined with structured conceptual synthesis, rather than as a systematic literature review. The purpose was not to provide exhaustive coverage of all research on generative AI in higher education, but to identify conceptually relevant sources capable of informing an institutional governance architecture for student research.

The source identification process focused primarily on literature published from 2022 to July 2026, reflecting the rapid development of generative AI applications in higher education during this period. Earlier sources were retained where they provided foundational concepts relating to academic integrity, AI literacy, trustworthy AI, or assessment. Literature identification drew on education-focused and open scholarly discovery sources, including ERIC-indexed

and publisher-hosted materials, supplemented by targeted scholarly discovery, citation tracing, and authoritative policy and guidance documents from international and national organisations.

The literature identification process was organised around five overlapping domains: (1) generative AI governance and institutional policy in higher education; (2) academic integrity, disclosure, transparency, authorship, and accountability; (3) assessment design and redesign in response to generative AI; (4) AI literacy, competency, implementation, and institutional support; and (5) generative AI use in student research, theses, dissertations, research projects, and research writing.

Sources were retained if they informed at least one of four areas: governance or integrity problems, relevant principles, institutional functions, or operational procedures. Eligible material comprised peer-reviewed empirical studies, conceptual papers, reviews, and authoritative policy or guidance documents. Work on technical model performance was excluded unless it had clear implications for higher education, student research, academic integrity, assessment, or institutional governance. Studies of attitudes or technology acceptance were included only when they offered substantive governance or implementation implications.

Viet Nam-specific policy and higher education sources were additionally considered to situate the framework within an illustrative emerging AI governance context. These sources were used to examine the relationship between broad national directions concerning AI and digital competence and the operational requirements of student research governance; they were not intended to constitute a comprehensive review of Vietnamese higher education policy.

3.2. Concept extraction and structured conceptual coding

Candidate sources identified through the targeted literature identification process were screened for conceptual relevance and examined using structured conceptual coding and mapping. The purpose was not to quantify the frequency of particular concepts or estimate effects, but to identify recurring and conceptually distinct governance problems and determine the institutional functions required to address them.

Concepts were organised into four analytical categories: (1) risks and governance problems, (2) normative principles, (3) institutional governance requirements, and (4) operational mechanisms or practices. These categories provided an initial deductive structure while allowing additional concepts to be incorporated when they were relevant to student research governance.

Recurring problems fell into three broad groups: unclear rules and inconsistent assessment; weak disclosure, unverified information, and over-reliance on AI outputs; and uneven capacity to interpret or meet institutional requirements. A related concern was whether assessment could identify students' independent scholarly contribution. These problems were examined in relation to principles drawn from academic integrity, responsible and trustworthy AI, AI literacy, and assessment scholarship, including transparency, accountability, fairness, human oversight, human agency, verification and reliability, privacy and data governance, and proportionality.

For each problem, the analysis identified the institutional requirement needed to put the relevant principle into practice. Inconsistent rules implied boundary setting and role clarification; undeclared AI use implied traceability; uncertainty about student contribution implied evaluability; incomplete disclosure implied assurance; and uneven capability implied capability building. Evidence that mandatory declarations do not ensure complete compliance supports the distinction between disclosure and assurance (Gonsalves, 2025). Fairness, privacy, human agency, human oversight, and proportionality were treated as cross-cutting principles.

3.3. Framework derivation and refinement

Framework development was abductive rather than wholly inductive. A provisional five-function architecture was derived from the governance problem and then examined against the identified risks, principles, requirements, and mechanisms; overlaps and gaps were used to refine the categories.

The mapping proceeded in three steps: linking problems to principles, translating principles into institutional requirements, and consolidating related requirements into the five functions shown in Table 2.

3.4. Operationalisation of the framework

Table 2 presents the resulting mapping and links each governance function to an illustrative operational mechanism. Section 5 considers how these mechanisms may be adapted in practice.

Table 2. Conceptual mapping from governance problems to framework functions and operational mechanisms

Governance problem or risk	Relevant principle(s)	Governance requirement	Governance function	Illustrative operational mechanism
Unclear or inconsistent rules concerning permissible AI use	Fairness; accountability	Boundary setting and role clarification	Policy	AI use policy template
Undeclared or inadequately declared AI assistance	Transparency	Traceability	Transparency	AI Transparency Appendix
Fabricated references or unverified AI-generated information	Reliability; human oversight; academic integrity	Verification	Assessment / Audit	Verification criteria and selective checks
Difficulty identifying students' independent scholarly contribution	Human agency; assessment validity	Evaluability	Assessment	AI-integrated rubric and process evidence
Over-reliance on AI-generated reasoning or content	Human agency; academic integrity	Demonstration of independent scholarly contribution	Assessment	Rubric criteria and process-based evidence
AI-use declarations may be incomplete or inaccurate	Accountability; proportionality	Assurance	Audit	Selective or risk-based audit
Inconsistent assessment expectations across supervisors or panels	Fairness; consistency	Shared assessment expectations	Policy / Assessment	Institutional guidance and common rubric
Students, supervisors, or assessors lack sufficient understanding	AI literacy; human oversight; fairness	Capability building	Guidance/support	Student guidance, staff orientation, examples, and checklists
Privacy, confidentiality, or sensitive-data risks	Privacy and data governance	Safe-use boundaries	Policy / Transparency / Guidance-support	Policy restrictions, approved-use rules, and disclosure guidance
Unequal access to AI tools or capabilities	Fairness; equity	Proportionate and equitable implementation	Policy / Assessment / Guidance-support	Access provisions, assessment adaptation, and support
Excessively burdensome or surveillance-like monitoring	Proportionality; fairness	Risk-based assurance	Audit	Selective rather than universal audit

The mapping clarifies that the five components are not isolated tools. Each responds to a distinct governance requirement, while some risks require more than one function. For example, fabricated references require routine verification through assessment but may also trigger selective audit, and inconsistent assessment expectations may require both policy-level standardisation and shared assessment criteria. This one-to-many mapping allows interaction among components without collapsing their distinct governance functions.

4. PROPOSED GOVERNANCE FRAMEWORK

The proposed framework is intended to help higher education institutions govern students' use of AI in research in ways that are transparent, verifiable, assessable, and accountable. It is not designed as a fixed regulation for one institution. Instead, it is an adaptable model that can be adjusted according to institutional context, disciplinary norms, technological infrastructure, academic regulations, and available support capacity.

The architecture combines a four-stage pathway - policy, transparency, assessment, and audit - with guidance/support as a cross-cutting function. Policy sets *ex ante* boundaries; transparency records process evidence; assessment interprets that evidence in relation to competence and academic standards; and audit provides selective assurance when further examination is warranted. Guidance/support enables implementation, while feedback from assessment and audit informs institutional learning. Figure 1 illustrates this logic, and Table 3 summarises the components and operational tools.

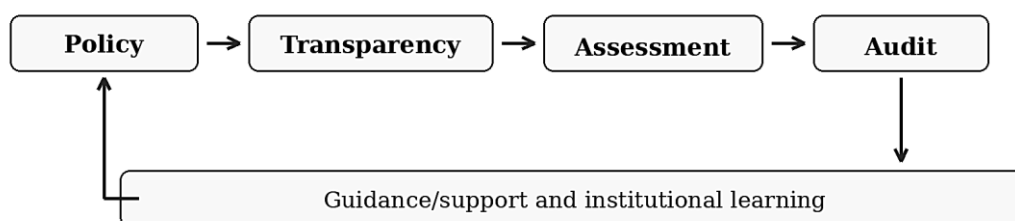


Figure 1. Structural logic of the proposed governance framework

Policy, transparency, assessment, and audit form a sequential governance pathway; guidance/support operates as a cross-cutting capability-building function, while feedback from implementation, assessment, and audit contributes to institutional learning and framework revision.

Table 3. Components, governance functions, and operational tools of the proposed framework

Component	Governance function	Primary governance question	Operational tool
Policy	Boundary setting	What AI uses are permitted, restricted, or prohibited, and who is responsible?	AI use policy template
Transparency	Traceability	Where, why, and how was AI used, and how was its output handled?	AI Transparency Appendix
Assessment	Evaluability	How should students' contribution and responsible AI use be evaluated?	AI-integrated rubric
Audit	Assurance	When and how should the credibility of disclosure and process evidence be further examined?	Selective or risk-based audit
Guidance/support	Capability building	How can students and staff understand and enact governance expectations?	Guidance, checklist, and orientation

4.1. Policy

The policy layer defines where students may use AI in research, which uses are prohibited, and who remains responsible for the final output. At the practical level, a policy may allow AI use for literature searching, summarisation, idea generation, drafting support, language checking, or preliminary suggestions for data handling.

It should prohibit fabricating data, generating false findings, creating non-existent references, copying AI-generated outputs without revision and verification, or allowing AI to replace students' own reasoning and analysis. This layer supports consistency among students, supervisors, and assessment panels.

4.2. Transparency

The transparency layer turns AI disclosure from a general statement into a traceable process. A sentence such as "AI was used for support" is insufficient because it does not explain where AI was used, why it was used, which outputs were retained, and how students verified or revised them. The framework therefore proposes an AI Transparency Appendix that records the tool used, purpose and scope of use, representative prompts, key outputs, student edits, risks or limitations, and verification sources. This appendix provides process evidence for supervision and assessment.

4.3. Assessment

The assessment layer asks how AI-supported work should be evaluated fairly. Assessment should not judge how well AI performs. It should assess students' academic competence when using AI: independent research thinking, disclosure, verification, revision and critique of AI-generated outputs, and academic ethics. An AI-integrated rubric can translate these expectations into criteria. Without such criteria, students may not know what counts as responsible AI use, and assessors may lack a shared basis for judgement.

4.4. Audit

The audit layer ensures that disclosure does not become merely performative. Disclosure creates process evidence, but evidence may occasionally require further assurance, particularly where declarations are incomplete, internally inconsistent, or raise specific integrity concerns. AI audits may involve asking students to explain their AI-use process, reproduce part of a prompt or workflow, compare the declaration with the final product, or explain how sources were verified. The framework favours selective or risk-based verification rather than routine auditing of every submission. This distinction is important because requiring AI-use declarations does not, by itself, guarantee complete compliance (Gonsalves, 2025). Audit should therefore be framed as a proportionate mechanism for accountability, fairness, and learning rather than primarily as punishment or universal surveillance.

4.5. Guidance and support

Policy, disclosure, and audit may be perceived as top-down control if relevant actors are not supported. Guidance should explain institutional expectations and show how to meet them through concise principles, checklists, examples of appropriate and inappropriate use, and instructions for disclosure and verification. Support should also extend beyond students: supervisors and assessors need sufficient AI competence to interpret policies, advise students, and apply assessment expectations consistently. UNESCO's AI Competency Framework for Teachers similarly emphasises human-centred, ethical, and continuing professional capability in relation to AI (Miao & Cukurova, 2024).

5. OPERATIONAL TOOLS AND ADAPTABILITY

The operational tools are designed for institutional adaptation. Their detail, mandatory status, disciplinary wording, submission procedures, and assessment weighting may vary, provided that the purpose of each governance function is preserved.

This adaptability is important because AI use in students' research projects differs across fields. A project in the social sciences or humanities may involve AI mainly for literature searching, summarisation, argument development, translation, or language support. A project in engineering, data science, or biomedical fields may involve AI in coding, data processing, simulation, visualisation, or preliminary analytical support. The same principles of transparency, verification, and responsibility therefore need to be specified differently across disciplines.

Supplementary File 1 contains an AI use policy template, an AI Transparency Appendix, an AI-integrated assessment rubric, a student guidance/checklist, and a selective audit procedure. These materials illustrate possible institutional procedures; they are not validated instruments or mandatory standards.

6. IMPLEMENTATION PATHWAY AND FUTURE EVALUATION

The framework should be introduced incrementally rather than implemented in full across an institution at once. A staged approach can reduce administrative burden and allow procedures to be refined before wider adoption.

Implementation may proceed through four phases. Preparation involves establishing or reviewing institutional AI policy and developing guidance for students and staff. Pilot implementation introduces transparency documentation and AI-integrated assessment criteria within selected programmes or research-project contexts. Assurance introduces proportionate, selective audit procedures after disclosure and assessment processes have been established. Institutional learning uses feedback from students, supervisors, assessors, and audit cases to revise policy, guidance, and assessment procedures.

Future evaluation should examine not only whether the framework is adopted, but whether it is feasible, reliable, equitable, and educationally defensible. At the student level, relevant indicators include disclosure completeness and the frequency of unverifiable claims or fabricated references. At the assessment level, evaluation may examine inter-rater consistency and assessor perceptions of rubric clarity. Governance-level indicators may include the frequency and outcomes of selective audits and the resources required for implementation. Institutional evaluation may examine adoption across departments, policy revision, and student and staff perceptions. These indicators are directions for future empirical evaluation rather than evidence of the framework's current effectiveness. Table 4 summarises illustrative indicators.

Table 4. Future evaluation indicators

Level	Suggested indicators	Purpose
Student	Completeness of AI disclosure; number of unverifiable claims or fabricated references	Transparency quality and integrity risk
Assessment	Inter-rater consistency; assessor feedback on rubric clarity	Fairness and assessment reliability
Governance	Proportion of submissions audited; discrepancies between declarations and outputs; implementation resources	Accountability, proportionality, and feasibility
Institutional	Policy revisions; departmental adoption; student and staff feedback	Scalability and organisational learning

7. DISCUSSION

7.1. From individual responsibility to institutional governability

The framework shifts responsible AI use from a purely individual ethical obligation to a distributed governance responsibility. Students remain accountable for their scholarly decisions and final submissions, while institutions establish the rules, documentation, assessment criteria, and support needed for consistent practice. This position builds on academic-integrity scholarship, which treats integrity as a shared educational and institutional responsibility, and on higher-education AI policy research that emphasises coordinated stakeholder roles (Macfarlane et al., 2014; Chan, 2023; An et al., 2025).

The framework also provides an alternative to both blanket prohibition and unrestricted use. Prohibition may discourage disclosure and is difficult to sustain as generative AI becomes embedded in common research and writing tools. Unrestricted use, however, can obscure students' independent contribution and weaken assessment validity. Recent scholarship therefore supports aligning permitted AI use with the purpose of the task, requiring transparent documentation, and redesigning assessment to make student competence visible (Luo, 2024; Perkins et al., 2024; Corbin et al., 2025).

7.2. Multi-level policy and implementation implications

Although the framework is primarily institutional, implementation depends on alignment across national, institutional, disciplinary, and project levels. This reflects evidence that broad policy principles require institutional procedures and locally interpretable guidance before they can shape teaching, assessment, and research practice (Chan, 2023; Moorhouse et al., 2023; An et al., 2025).

At the national or system level, authorities can establish broad expectations for responsible AI use, academic integrity, digital competence, data protection, and educational quality. In Viet Nam, current strategies and competency initiatives provide this enabling direction, but they do not prescribe detailed procedures for individual student research projects (Prime Minister of Viet Nam, 2021; Ministry of Education and Training, 2025a, 2025b).

At the institutional level, universities translate broad principles into operational rules for permitted use, disclosure, assessment, verification, and support. Reviews of university AI policies show that institutional guidance increasingly addresses these areas, although their coverage across topics and stakeholder groups remains uneven (Luo, 2024; An et al., 2025). The institution is therefore the main level at which the five governance functions can be coordinated.

At the faculty or department level, common institutional expectations require disciplinary adaptation. The implications of AI-supported coding, language editing, literature exploration, or image generation differ across fields and assessment formats. Assessment scholarship similarly argues that acceptable AI use must be aligned with the capabilities a task is intended to evidence (Perkins et al., 2024; Corbin et al., 2025).

At the supervisor level, governance becomes project-specific and educational. Supervisors can clarify acceptable uses, support verification and disclosure, and help students distinguish assistance from substitution of scholarly reasoning. Their role should nevertheless operate within shared institutional rules, because reliance on individual judgement alone can produce inconsistent expectations (Chan, 2023; An et al., 2025).

These levels are complementary rather than interchangeable. National direction cannot substitute for institutional procedures; institutional rules require disciplinary interpretation; and supervisors need common expectations. In Viet Nam, such alignment could connect emerging national priorities with the practical concerns about AI literacy and academic integrity already reported among students and educators (Ngo et al., 2024; Le et al., 2025; Do et al., 2026).

7.3. Governance tensions in implementation

Responsible AI governance involves tensions that rules alone cannot remove. The first concerns transparency and assessment fairness. Disclosure can improve traceability, but students who report AI use honestly should not be disadvantaged merely because they disclosed it. Research on AI-use declarations also shows that a formal requirement does not guarantee complete compliance (Gonsalves, 2025). Assessment should therefore evaluate appropriateness, verification, critical judgement, and scholarly contribution rather than treating AI use itself as misconduct.

A second tension concerns standardisation and disciplinary autonomy. Shared expectations can improve fairness, yet uniform rules may not reflect differences in research practice and assessment purpose. Existing assessment frameworks support explicit but context-sensitive decisions about the role of AI in a task (Perkins et al., 2024; Luo, 2024). The practical objective is therefore to standardise core governance functions while allowing disciplines to adapt boundaries and examples.

A third tension concerns accountability and surveillance. Selective verification can protect assessment fairness, but excessive monitoring may create burden, undermine trust, and encourage defensive compliance. This concern is especially relevant because automated inference or detection cannot substitute for evidence about a student's research process. A proportionate audit should therefore respond to defined risks or inconsistencies and give students an opportunity to explain their work (Gonsalves, 2025; Miao & Holmes, 2023).

These tensions support treating governance as an iterative institutional-learning process. Assessment experience, student and staff feedback, and selective audit can reveal weaknesses in policy, guidance, or assessor preparation. This approach is consistent with responsible-AI and competency frameworks that emphasise continuing human oversight, capability development, and adaptation as technologies and contexts change (Miao & Cukurova, 2024; Miao & Holmes, 2023; Miao et al., 2024).

8. LIMITATIONS AND FUTURE RESEARCH

This paper has several limitations. First, the proposed framework is conceptual and design-oriented and has not yet been empirically evaluated. Its effectiveness in improving disclosure, source verification, assessment consistency, or academic integrity therefore remains to be established.

Second, the targeted literature identification was designed to strengthen conceptual relevance and traceability rather than to provide exhaustive systematic coverage. Relevant literature may therefore have been missed, and the framework should be revisited as the evidence base and institutional practices continue to develop.

Third, the associated operational tools - including the AI use policy template, AI Transparency Appendix, assessment rubric, audit procedure, and guidance materials - are design artefacts rather than validated instruments. Their feasibility, usability, reliability, and disciplinary fit require empirical testing.

Fourth, Viet Nam is used as an illustrative emerging AI governance context rather than as an empirical national case study. The framework should therefore not be interpreted as a comprehensive representation of Vietnamese higher education policy or institutional practice.

Finally, implementation may generate unintended burdens or inequities. Detailed disclosure requirements may increase administrative workload, audit may be perceived as surveillance, and differences in access to AI tools may affect assessment fairness. Future studies should therefore examine not only effectiveness but also feasibility, acceptability, proportionality, equity, and transferability across disciplines and institutional settings.

Priority areas for future research include piloting the AI Transparency Appendix, examining the inter-rater reliability of the AI-integrated rubric, evaluating student and supervisor experiences of selective audit, and testing whether the framework improves disclosure quality, source verification, assessment consistency, and institutional capability.

9. CONCLUSION

Generative AI is becoming embedded in the processes through which students search, analyse, draft, and communicate research. The central governance challenge is therefore not simply whether AI should be permitted, but how its use can remain compatible with student responsibility, academic integrity, and meaningful assessment.

This paper proposed a five-function governance architecture for students' research projects, connecting boundary setting, traceability, evaluability, assurance, and capability building. Its contribution lies not in presenting these mechanisms as individually novel, but in showing how they can operate as an integrated institutional governance process linking supervision, documentation, assessment, review, and support.

The framework remains conceptual and requires empirical evaluation. Its operational tools should be adapted to disciplinary and institutional contexts and implemented proportionately. In emerging AI governance contexts such as Viet Nam, the framework may provide one approach for connecting national AI and digital-competence priorities with institutional procedures, disciplinary adaptation, and project-level supervision.

Future research should examine whether the framework improves disclosure quality, verification practices, assessment consistency, and institutional capability without creating excessive administrative burden or surveillance. Responsible AI governance in student research should ultimately preserve not only academic integrity, but also students' independent judgement, scholarly responsibility, and human agency in knowledge production.

Conflict of Interest: No potential conflict of interest relevant to this article was reported.

REFERENCES

- An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22, Article 10. <https://doi.org/10.1186/s41239-025-00507-3>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, Article 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, 50(7), 1087-1097. <https://doi.org/10.1080/02602938.2025.2503964>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- Do, H. V., Bui, T. T., Tran, D. H., Nguyen, H. T., & Dinh, L. D. (2026). Developing and validating an AI literacy scale for university students in Viet Nam. *IFLA Journal*, 52(2), 222-233. <https://doi.org/10.1177/03400352251385058>
- Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity*, 19, Article 23. <https://doi.org/10.1007/s40979-023-00144-1>

- European Commission 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>
- Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 50(4), 592-606. <https://doi.org/10.1080/02602938.2024.2415654>
- Le, T. P., Do, D. K., Hoang, H. D., & Nguyen, X. H. (2025). AI tools as facilitators or threats to Vietnamese students' academic integrity in research writing. *AsiaCALL Online Journal*, 16(2), 41-55. <https://doi.org/10.54855/acoj.251623>
- 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 598, pp. 1-16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the originality of students' work. *Assessment & Evaluation in Higher Education*, 49(5), 651-664. <https://doi.org/10.1080/02602938.2024.2309963>
- Macfarlane, B., Zhang, J., & Pun, A. (2014). Academic integrity: A review of the literature. *Studies in Higher Education*, 39(2), 339-358. <https://doi.org/10.1080/03075079.2012.709495>
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. <https://doi.org/10.54675/ZJTE2084>
- 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>
- Ministry of Education and Training (2025a). *Quyết định số 1504/QĐ-BGDĐT ngày 27/5/2025 ban hành Chương trình phổ cập kiến thức, kỹ năng số cho sinh viên trong các cơ sở giáo dục đại học* [Decision No. 1504/QĐ-BGDĐT dated May 27, 2025 promulgating the digital literacy and skills development programme for students in higher education institutions].
- Ministry of Education and Training (2025b). *Thông tư số 02/2025/TT-BGDĐT ngày 24/01/2025 quy định Khung năng lực số cho người học* [Circular No. 02/2025/TT-BGDĐT dated January 24, 2025 regulating the digital competency framework for learners].
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, Article 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ngo, C.-L., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology*, 24(1), 28-47. <https://doi.org/10.56297/FSYB3031/MXNB7567>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching & Learning Practice*, 21(6), Article 6. <https://doi.org/10.53761/q3azde36>
- Prime Minister of Viet Nam. (2021). *Quyết định số 127/QĐ-TTg ngày 26/01/2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030* [Decision No. 127/QĐ-TTg dated January 26, 2021 promulgating the national strategy on research, development, and application of artificial intelligence through 2030].
- UNICEF Viet Nam (2025, May 30). *Viet Nam National Forum on Artificial Intelligence in Education* [Press release] <https://www.unicef.org/vietnam/press-releases/viet-nam-national-forum-artificial-intelligence-education>