Available on: https://jmrr.ep-journals.org/ E-ISSN: 3051-3588, P-ISSN: 3051-357X

Journal of Management Research and Review

Volume: 02 Issue: 10 October-2026

Responsible Governance of Artificial Intelligence in the Design and Delivery of Higher Education Curricula in Vietnam

Nguyễn Văn Thản

Department of Business – Commerce, Faculty of Economics, Hanoi Industrial and Trade University, Vietnam.

Published Date: October 29, 2026 **Article DOI: 10.65150/EP-jmrr/V2E10/2026-07**

ABSTRACT:

Purpose: This study examines the risks that artificial intelligence (AI) creates across the curriculum lifecycle in higher education and identifies the governance mechanisms required for its responsible adoption in Vietnam.

Design/methodology/approach: The study combines a literature review, qualitative document analysis, and comparative policy benchmarking of the UNESCO Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, the European Union Artificial Intelligence Act, and Vietnam's Law on Artificial Intelligence, Law on Personal Data Protection, and National Artificial Intelligence Ethics Framework.

Findings: AI can support needs analysis, learning outcome mapping, personalized learning pathways, instructional delivery, and assessment. However, its use gives rise to six groups of risks: personal data violations, algorithmic bias, lack of transparency and explicability, diffused accountability, unequal access, and the erosion of academic autonomy. Vietnam established a general legal and ethical foundation for AI governance in 2025–2026; the main remaining gap lies in translating these general provisions into operational procedures specific to higher education.

Practical implications: The study proposes a ten-step, lifecycle-based governance framework comprising problem definition, risk classification, data mapping, impact assessment, pedagogical validation, technical and fairness testing, controlled piloting, accountable approval, post-deployment monitoring, and audit or discontinuation, together with a tiered allocation of responsibilities and the evidence required at each step.

Originality/value: The framework links international principles and Vietnam's newly enacted legal requirements to the concrete stages of curriculum design and delivery, providing higher education institutions and policymakers with a practical basis for transparent, fair, and accountable AI adoption that preserves academic integrity and learner autonomy.

KEYWORDS: Artificial intelligence, AI governance, responsible AI, higher education, curriculum, data protection, Vietnam.

1. INTRODUCTION

Artificial intelligence (AI) is increasingly used in higher education to analyze admissions data, academic performance, and labor market demand, to support the formulation of intended learning outcomes, to review course modules, and to personalize learning pathways (Zawacki-Richter et al., 2019). Virtual teaching assistants, conversational chatbots, and adaptive learning platforms extend the support available to students beyond the classroom, and the rapid spread of generative AI since 2022 has further widened the range of possible uses in teaching, learning, and assessment (UNESCO, 2023).

These applications, however, also raise concerns about privacy, algorithmic bias, opacity, and diffused accountability (Holmes et al., 2022; Slade & Prinsloo, 2013). In education, such concerns carry particular weight because AI-supported decisions may affect admission, academic progression, assessment results, and, ultimately, learners' future opportunities. The technical efficiency of AI therefore needs to be evaluated together with its pedagogical value, its effects on equity, and its implications for learner rights.

In Vietnam, the Law on Personal Data Protection took effect on 1 January 2026, the Law on Artificial Intelligence took effect on 1 March 2026, and the National Artificial Intelligence Ethics Framework was issued under Circular No. 05/2026/TT-BKHCN (Ministry of Science and Technology, 2026; National Assembly, 2025a, 2025b). Together, these instruments establish an initial legal and ethical foundation for responsible AI. Their provisions, however, are general and cross-sectoral, and higher education institutions still lack specific guidance on how to apply them to admissions, learning analytics, automated grading, and online examination proctoring. Although international studies have examined the ethics of AI in education, relatively few have connected these principles with the stages of the curriculum lifecycle and with Vietnam's newly enacted regulations.

To address this gap, this study investigates two research questions: (RQ1) What systemic risks does AI create across the processes of curriculum design and delivery in higher education? (RQ2) What governance mechanisms are needed to ensure the responsible adoption of AI in Vietnamese higher education? The study makes three contributions. First, it maps AI applications and their associated risks across the curriculum lifecycle.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Second, it benchmarks Vietnam's regulatory framework against those of UNESCO, the OECD, and the European Union to identify policy gaps. Third, it proposes a ten-step, lifecycle-based governance framework that specifies the responsibilities of each stakeholder and the evidence to be produced at each step, while ensuring that humans retain final decision-making authority. The remainder of the paper is organized as follows: Section 2 reviews the literature and presents the analytical framework; Section 3 describes the methodology; Section 4 presents and discusses the results; and Section 5 concludes with recommendations, limitations, and directions for future research.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 AI in higher education curricula

In higher education, AI refers to algorithmic systems capable of processing large datasets, performing automated reasoning, or generating content to support teaching, learning, assessment, and institutional administration. In this study, a curriculum is understood not merely as a list of course modules but as an integrated pedagogical system comprising educational objectives, intended learning outcomes (ILOs), content, teaching and learning methods, assessment, and mechanisms for continuous quality improvement. This view is consistent with Vietnamese regulations on training program standards, which require programs to be designed around learning outcomes and to be periodically evaluated and updated (Ministry of Education and Training, 2021).

In a systematic review of research on AI in higher education, Zawacki-Richter et al. (2019) identified four main areas of application: profiling and prediction, assessment and evaluation, adaptive systems and personalization, and intelligent tutoring systems. The authors also noted that most studies were technology-driven, rarely engaged with pedagogical theory, and paid little critical attention to ethical risks. This observation remains relevant, as institutions often adopt AI tools for specific tasks without considering their effects on the curriculum as a whole.

Because assessment shapes how teachers teach and how students learn, a phenomenon known as washback (Alderson & Wall, 1993), the introduction of AI even into a single function such as grading can have system-wide effects. It may change how lecturers design learning tasks and how students choose their learning strategies. AI governance in higher education should therefore be considered at the level of the curriculum system rather than tool by tool.

2.2 Responsible AI and governance principles

Responsible AI refers to the development and use of AI in ways that maximize benefits while systematically identifying, mitigating, and controlling adverse effects. A review of 84 AI ethics guidelines by Jobin, Ienca, and Vayena (2019) found convergence around five principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy. Synthesizing earlier frameworks, Floridi and Cowls (2019) proposed five principles for AI in society: beneficence, non-maleficence, autonomy, justice, and explicability. These principles have informed major international instruments, including the UNESCO (2022) Recommendation on the Ethics of Artificial Intelligence and the OECD (2024) AI Principles.

Principles alone, however, do not guarantee ethical practice. Mittelstadt (2019) argues that high-level principles are abstract, lack enforcement mechanisms, and can be interpreted in conflicting ways. Governance therefore requires translating principles into procedures, roles, and verifiable evidence. Raji et al. (2020) propose an end-to-end internal audit framework in which each stage of the AI lifecycle produces documentation that can be reviewed, while the NIST (2023) AI Risk Management Framework and the ISO/IEC 42001:2023 standard organize AI governance as a continuous management cycle of mapping, measuring, and managing risks.

In education, Holmes et al. (2022) argue that the ethics of AI must address not only data and algorithms but also pedagogy, that is, whether a system supports sound teaching and learning. Baker and Hawn (2022) show that algorithmic bias in educational systems can disadvantage learners by gender, ethnicity, socioeconomic status, and disability, while Slade and Prinsloo (2013) frame learning analytics as a moral practice shaped by the power relationship between institutions and students.

Building on this literature, this study operationalizes responsible AI through a three-tier governance structure: (1) a normative tier, which defines fundamental values, learner rights, and pedagogical safeguards; (2) an institutional tier, which establishes mandates, decision-making authority, ethics review procedures, and oversight processes; and (3) an operational tier, which produces verifiable evidence of compliance, such as data lineage records, test results, decision logs, bias audits, and records of corrective actions. This structure helps avoid two common errors: treating ethics as a set of declarations disconnected from practice, and equating technical compliance with comprehensive governance. A system can satisfy formal technical requirements while still undermining students' intellectual autonomy or reducing pedagogical depth. Effective governance must therefore balance pedagogical effectiveness, legal compliance, and social accountability. For high-stakes decisions concerning admission, academic progression, formal assessment, or degree conferral, human educators must retain final authority and the obligation to explain outcomes. Figure 1 summarizes the analytical framework of the study.

Figure 1. Analytical framework of the relationship between AI applications, ethical risks, and educational outcomes

AI applications	Operational impact	Emerging risks	Governance controls	Educational goals
Curriculum design, delivery, assessment & administration	Analytics, personalization, feedback & automation	Data privacy, algorithmic bias & system opacity	Human oversight, auditing & validation protocols	Quality, equity, safety & institutional trust

(Source: Compiled by the author from Floridi and Cowls (2019), UNESCO (2022, 2023), and OECD (2024))

3. RESEARCH METHODOLOGY

This study adopts a qualitative design that combines three methods: a literature review (Snyder, 2019), qualitative document analysis (Bowen, 2009), and comparative policy benchmarking. This design is appropriate because the research questions concern the conceptual mapping of risks and the design of governance mechanisms rather than the measurement of statistical relationships.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Two groups of sources were analyzed. The first group comprises peer-reviewed literature on AI applications in higher education, the ethics of AI in education, learning analytics, and algorithmic accountability. Sources were selected for their direct relevance to curriculum design, delivery, and assessment, or to AI governance, and for their influence in the field. The second group comprises official policy and regulatory documents: the UNESCO (2022) Recommendation on the Ethics of Artificial Intelligence; the UNESCO (2023) guidance on generative AI in education and research; the OECD (2024) Recommendation on Artificial Intelligence; Regulation (EU) 2024/1689 (the Artificial Intelligence Act); the NIST (2023) AI Risk Management Framework; and Vietnam's Law on Personal Data Protection, Law on Artificial Intelligence, and National Artificial Intelligence Ethics Framework (Ministry of Science and Technology, 2026; National Assembly, 2025a, 2025b). Documents were included if they were official, in force or formally adopted, and directly addressed AI governance or personal data protection.

The analysis proceeded in four steps. First, AI applications were mapped to five stages of the curriculum lifecycle: needs analysis, structural design, personalization, instructional delivery, and assessment. Second, the risks associated with each application were identified and coded into risk categories, using the principles of Floridi and Cowls (2019) as the coding frame. Third, the governance provisions of the selected frameworks were compared across five components (foundation, risk management, data, transparency, and oversight) to identify gaps in Vietnam's current framework as applied to higher education. Fourth, the findings were synthesized into a lifecycle-based governance framework, drawing on the lifecycle approaches of the OECD (2024), NIST (2023), and Raji et al. (2020). As a conceptual and normative study, the proposed framework has not yet been empirically tested; this limitation is discussed in Section 5.3.

4. RESULTS AND DISCUSSION

4.1 AI applications in curriculum design

In the needs analysis phase, AI can synthesize data on admissions, academic performance, graduate employment, employer feedback, and skill trends. Natural language processing models can analyze job advertisements and occupational standards at scale. These results do not replace consultation with experts and employers, but they can help identify emerging skills, redundant content, and gaps between learning outcomes and labor market requirements. The main value of AI at this stage lies in broadening the evidence base and shortening the review cycle.

In structural design, AI can map learning outcomes to modules, learning activities, and assessment methods, and can detect learning outcomes that are not assessed, modules that are overloaded, or competencies that are unevenly distributed across the program. Within outcome-based education (OBE), CDIO, or AUN-QA frameworks, AI can help organize evidence and track improvements. However, keyword matching does not guarantee pedagogical appropriateness. Any proposed change to the curriculum must be reviewed by the relevant academic committee in light of the discipline, available resources, and learner characteristics.

AI can also support the design of flexible learning pathways. On the basis of completed modules, prior competencies, and career goals, a system can suggest elective modules, supplementary materials, or remediation plans. Such personalization is beneficial only if it does not lower academic standards or confine learners to paths predicted from their past performance. Students should be informed of the basis for recommendations, be able to choose alternative pathways, and have access to academic advising. Table 1 summarizes the main categories of AI application, together with their benefits, risks, and corresponding control measures.

Table 1. Categories of AI application in higher education curricula and associated risks

Phase	AI application	Benefits	Risks	Control measures
Needs analysis	Analysis of employment data, stakeholder feedback, and learning outcomes	Identification of emerging skill gaps	Biased datasets; flawed statistical inferences	Representativeness checks; expert validation
Structural design	Mapping learning outcomes to modules and assessments	Detection of curricular gaps and redundancies	Equating keyword matching with pedagogical quality	Review and approval by the academic committee
Personalization	Suggesting elective modules, materials, and pathways	Targeted support tailored to student needs	Confining learners to predicted pathways	Explanation of recommendations; right to decline
Instructional delivery	Chatbots, virtual assistants, adaptive learning	Support outside class hours; continuous feedback	Inaccurate outputs (hallucinations); reduced human interaction	Content verification; instructor oversight
Assessment	Grading structured items, error diagnosis, rubric-based feedback	Reduced workload; immediate feedback	Difficulty explaining scores; grading bias	Defined scope of automation; right to human regrading

(Source: Synthesized by the author from Zawacki-Richter et al. (2019), Slade & Prinsloo (2013), and UNESCO (2023))

4.2 AI applications in curriculum delivery and assessment

In instructional delivery, virtual teaching assistants and academic chatbots can answer frequently asked questions, guide students in the use of learning materials, and support review outside class hours. Adaptive learning systems adjust the difficulty, sequencing, or type of exercises according to learner performance, while learning analytics help instructors monitor engagement and progress and identify students who may need support. These tools are particularly useful in large classes, but they are effective only when their content is verified, their limitations are disclosed, and instructors are able to correct system errors.

In assessment, AI can grade structured questions, classify feedback, detect patterns of errors, and generate preliminary comments. For essays and other open-ended work, AI should serve as a support tool rather than as the final grader. The reliability of automated feedback depends on the

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

training data, the grading rubric, and the system's ability to interpret context; fluent feedback may still be inaccurate or misaligned with module objectives. Institutions should therefore specify which assessment tasks may be automated, the thresholds at which human verification is required, and the procedures through which students can request regrading.

Generative AI further expands the possibilities for creating examples, simulating scenarios, suggesting questions, and supporting academic writing. At the same time, it blurs the boundary between legitimate assistance and outsourcing of intellectual work. UNESCO (2023) emphasizes that AI outputs are not authoritative sources of knowledge and must be verified by users. Assessment should therefore shift from evaluating only the final product toward evaluating the learning process, the evidence underlying decisions, the ability to defend arguments, and transparency about the extent of AI use. The aim is not outright prohibition but the protection of cognitive engagement, academic integrity, and individual responsibility.

4.3 Ethical and legal risks

The analysis identified six groups of risks. The first concerns learner data. AI systems may process grades, access logs, time spent on tasks, interaction content, and inferred competency data. When learners do not know what data are collected, for what purpose, and for how long they are retained, consent easily becomes a formality. Slade and Prinsloo (2013) situate learning analytics within the power relationship between institutions and students and highlight issues of consent, privacy, transparency, and data access rights. Under Vietnam's current legal framework, these safeguards must be complemented by purpose limitation, data minimization, access control, audit trails, and the protection of data subjects' rights (Ministry of Science and Technology, 2026; National Assembly, 2025a).

The second group of risks concerns bias. Historical data may reflect disparities related to region, socioeconomic status, gender, disability, or access to technology. When such data are used to predict academic success, a system may convert past inequalities into present recommendations (Baker & Hawn, 2022). Bias can also arise when the optimization criteria of a system are misaligned with educational goals, for example when completion speed is treated as the sole indicator of competence. Controlling bias requires assessing data representativeness, comparing error rates across learner groups, consulting affected groups, and monitoring the system after deployment.

The third group concerns the lack of transparency and explicability. Learners may receive grades, warnings, or recommendations without knowing which factors produced them. For high-stakes decisions, a general description of the algorithm is not a sufficient explanation. Institutions should provide enough information for learners to understand the purpose of the system, the main data used, its limitations, the role of human decision-makers, and the procedure for requesting a review.

The fourth group concerns the accountability gap. Institutions often procure AI services from vendors, which in turn may rely on third-party models. When a system fails, responsibility can easily be shifted between parties (Raji et al., 2020). Contracts should therefore specify audit rights, incident notification, data retention and transfer, responsibility for correcting errors, and conditions for termination. Although institutions may outsource technology, they cannot outsource academic responsibility.

The fifth group concerns unequal access. Paid tools, personal devices, and the quality of internet connections create disparities in students' ability to benefit from AI. Students with disabilities and those who use low-resource languages may receive lower-quality services. Institutional policies should provide equivalent alternatives, meet accessibility standards, and offer technical support.

The sixth group concerns autonomy and academic culture. Excessive reliance on automated suggestions can narrow pedagogical choices and reduce opportunities for independent thinking, while pervasive automated monitoring may create a sense of constant surveillance. Responsible governance should identify tasks that should not be fully delegated to machines, preserve human interaction in teaching and learning, and assess whether a technology genuinely improves learning or merely increases measurement capacity. These six groups of risks are interrelated: unrepresentative data can produce biased outcomes, opaque models limit learners' ability to contest decisions, and ambiguous contracts can shift responsibility away from the institution. Risk control must therefore be applied throughout the system lifecycle rather than at a single point.

4.4 Policy benchmarking and gaps in Vietnam

International frameworks reflect three complementary approaches. UNESCO (2022) places human dignity, human rights, and ethical impact assessment at the center of AI governance. The OECD (2024) emphasizes trustworthy AI that is safe, transparent, explainable, and accountable throughout its lifecycle. The European Union translates these principles into legally binding obligations calibrated to risk levels. Under Annex III of Regulation (EU) 2024/1689, AI systems intended to determine access or admission to educational institutions, evaluate learning outcomes, assess the appropriate level of education, or monitor prohibited behavior during tests are classified as high-risk (European Parliament & Council of the European Union, 2024).

Vietnam has recently established a general legal foundation for AI governance. The Law on Personal Data Protection (No. 91/2025/QH15) took effect on 1 January 2026, and the Law on Artificial Intelligence (No. 134/2025/QH15), adopted on 10 December 2025, took effect on 1 March 2026 (National Assembly, 2025a, 2025b). The Law on Artificial Intelligence classifies AI systems into high, medium, and low risk levels according to their potential impact on human rights, safety, security, and the public interest, and requires high-risk systems to undergo conformity assessment before deployment. The National Artificial Intelligence Ethics Framework, effective from 10 March 2026, sets out principles of safety and reliability with human oversight; respect for human rights, fairness, transparency, the limitation of data bias, privacy protection, and clear accountability; social benefit; and responsible innovation (Ministry of Science and Technology, 2026).

The main gap therefore lies not in the absence of principles but in the lack of operational standards specific to higher education. Six areas require further specification: (1) risk classification by educational use case; (2) pre-deployment impact assessment; (3) evaluation of pedagogical effectiveness; (4) requirements for procurement and vendor management; (5) mechanisms for explanation, appeal, and regrading; and (6) independent auditing of systems that affect academic rights. In addition, guidance is needed on the governance of learning analytics data, particularly data inferred about learners' competencies or dropout risk. Table 2 summarizes the comparison.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Table 2. Benchmarking of AI governance components relevant to education

Component	Vietnam	UNESCO / OECD	European Union (AI Act)	Implications for higher education
Foundation	Safety and reliability; human rights, fairness, and accountability; social benefit; responsible innovation	Human-centered, trustworthy AI	Protection of fundamental rights	Testable pedagogical criteria and an ethical baseline
Risks	Three risk levels (high, medium, low); not specific to education	Risk and impact assessment across the lifecycle	Specified educational uses classified as high-risk	Inventory of use cases and risk thresholds
Data	Law on Personal Data Protection in force since 1 January 2026	Purpose limitation and privacy protection	Strict data governance for high-risk systems	Guidance on learning analytics and inferred data
Transparency	Transparency and clear accountability	Explicability and accountability	Binding obligations for providers and deployers	Notification templates and explanation dossiers
Oversight	Human oversight; conformity assessment of high-risk systems	Impact assessment and stakeholder consultation	Risk management, logging, and post-market monitoring	Regrading rights, independent audits, and suspension rules

(Source: Synthesized by the author from National Assembly (2025a, 2025b), Ministry of Science and Technology (2026), UNESCO (2022), OECD (2024), and European Parliament & Council of the European Union (2024))

4.5 Proposed lifecycle-based AI governance framework

Based on the preceding analysis, this study proposes a ten-step governance framework organized into four stages of the AI lifecycle (Figure 2).

Planning (Steps 1–3). The institution first defines the educational problem to be solved and considers whether AI is necessary or whether an existing solution is sufficient (Step 1). The system is then classified by risk level according to the nature of the decisions involved, the number of people affected, the sensitivity of the data, and the reversibility of potential errors (Step 2). The institution next maps the data, specifying their sources, purposes of use, retention periods, and access rights (Step 3).

Assessment (Steps 4–6). The ethical, legal, and human rights impacts of the system are assessed (Step 4). Its pedagogical value is then validated: effectiveness should be judged primarily by its contribution to learning quality rather than by processing speed or cost savings (Step 5). The system is subsequently tested for accuracy, robustness, and differences in performance across learner groups (Step 6).

Deployment (Steps 7–8). Deployment begins with a controlled pilot that has clearly defined boundaries, is fully disclosed to stakeholders, and retains non-AI alternatives so that learners can choose (Step 7). Before full rollout, the system is approved on the basis of an accountability dossier that identifies the final decision-maker, the contact point for complaints, and the conditions for suspension or termination (Step 8).

Operation (Steps 9–10). After deployment, the system is continuously monitored using indicators of effectiveness, bias, incidents, user feedback, and impacts on vulnerable groups (Step 9). Finally, the system is periodically audited and then improved or discontinued as necessary (Step 10). Prior investment or existing contracts cannot justify the continued use of a technology whose benefits no longer outweigh its risks. The results of Step 10 feed back into Step 1, making the framework a continuous improvement cycle.

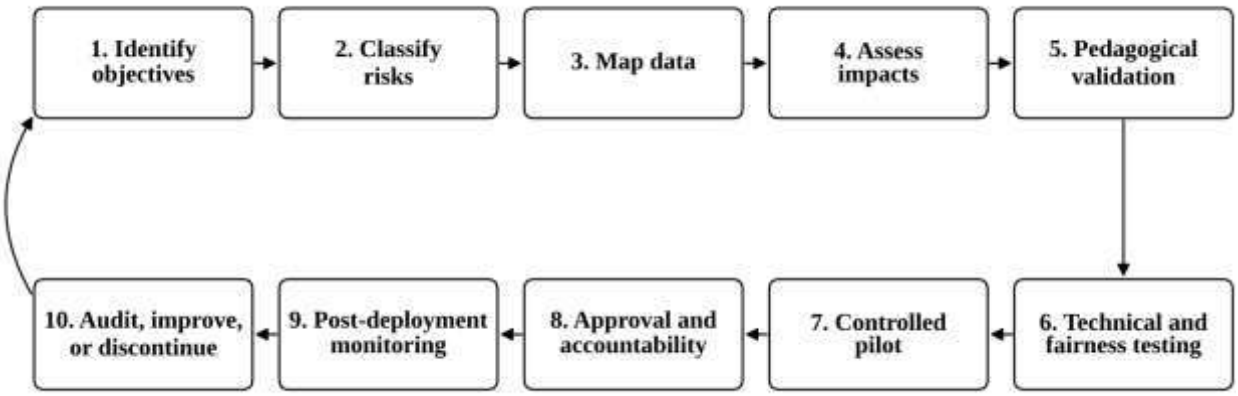


Figure 2. Ten-step lifecycle-based AI governance framework for higher education
(Source: Proposed by the author)

Responsibilities within the framework are tiered. Regulatory authorities issue minimum criteria and supervise high-risk cases. Higher education institutions develop internal policies, establish AI review boards, and maintain inventories of AI systems. Academic units define pedagogical objectives and remain accountable for academic decisions. Information technology, legal, and data protection units evaluate safety, contracts, and compliance. Providers supply technical documentation, logs, information on system limitations, and audit support. Instructors verify system outputs, while students are informed about AI use and retain the right to appeal. Each step produces verifiable evidence, as summarized in Table 3. In this way, ethics moves beyond declarations and generates evidence that can also inform curriculum improvement.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Table 3. Key questions, evidence, and responsibilities in the ten-step governance framework

Stage	Step	Key question	Evidence produced	Lead responsibility
Planning	1. Identify objectives	What educational problem is addressed, and is AI necessary?	Problem statement and needs justification	Academic unit
Planning	2. Classify risks	How severe and reversible are the potential harms?	Risk classification form	AI review board
Planning	3. Map data	Which data are used, for what purpose, for how long, and by whom?	Data inventory and data-flow map	Data protection unit
Assessment	4. Assess impacts	What are the ethical, legal, and human rights impacts?	Impact assessment report	AI review board; legal unit
Assessment	5. Pedagogical validation	Does the system improve learning quality?	Pedagogical evaluation report	Academic committee
Assessment	6. Technical and fairness testing	Is the system accurate, robust, and fair across learner groups?	Test results disaggregated by learner group	IT unit; provider
Deployment	7. Controlled pilot	Does the system work under real conditions with safeguards?	Pilot protocol, disclosure notice, non-AI alternative	Academic unit; instructors
Deployment	8. Approval and accountability	Who decides, who handles complaints, and when should use stop?	Accountability dossier and approval record	Institutional leadership
Operation	9. Post-deployment monitoring	Do effectiveness, bias, and incidents remain within thresholds?	Monitoring logs and incident reports	IT unit; quality assurance unit
Operation	10. Audit, improve, or discontinue	Do the benefits still outweigh the risks?	Audit report; remediation or discontinuation decision	Independent auditor; leadership

(Source: Proposed by the author based on OECD (2024), NIST (2023), and Raji et al. (2020))

4.6 Discussion

The findings for RQ1 show that the risks of AI in higher education are not confined to a single tool or decision but are distributed across the entire curriculum lifecycle, from needs analysis to assessment. This finding is consistent with the observation of Zawacki-Richter et al. (2019) that research on AI in higher education has paid insufficient attention to pedagogical and ethical issues, and with the argument of Holmes et al. (2022) that the ethics of AI in education must encompass pedagogy as well as data and algorithms. Because of washback effects, risks arising in assessment can propagate to teaching and learning practices, which reinforces the need for system-level governance.

Regarding RQ2, the proposed framework shares the lifecycle logic of the OECD (2024) AI Principles, the NIST (2023) AI Risk Management Framework, and ISO/IEC 42001:2023, but adapts it to higher education in two respects. First, it introduces a separate step of pedagogical validation, so that AI systems are evaluated against learning quality rather than technical performance alone. Second, it links each step to specific evidence and responsible actors, responding to the critique that ethical principles lack mechanisms for implementation (Mittelstadt, 2019) and reflecting the documentation-based audit approach of Raji et al. (2020).

For Vietnam, the framework also offers a way to operationalize the newly enacted Law on Artificial Intelligence and National Artificial Intelligence Ethics Framework within individual institutions while sector-specific guidance is being developed. Given the uneven resources of Vietnamese higher education institutions, the framework should be applied proportionately: high-risk use cases such as admissions, high-stakes assessment, and online proctoring should be subject to all ten steps, whereas low-risk applications, such as tools that generate practice exercises, may follow a simplified procedure.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study shows that AI has been integrated into many stages of curriculum design and delivery in higher education, including needs analysis, learning outcome mapping, curriculum review, learning pathway recommendations, progress monitoring, assessment, and feedback. The appropriate role of AI, however, is to broaden the evidence base and support decision-making. Decisions concerning educational objectives, learning outcomes, academic results, and learner rights must remain with academic committees, instructors, and institutions.

In response to RQ1, the study identifies six interrelated groups of risks: personal data violations, algorithmic bias, lack of transparency and explicability, diffused accountability, unequal access, and the erosion of academic autonomy. In response to RQ2, the study finds that Vietnam established an important legal and ethical foundation for AI governance in 2025–2026, but that the main challenge is to translate general principles into procedures specific to higher education. The proposed ten-step framework addresses this challenge through risk classification, impact assessment, pedagogical validation, human oversight, continuous monitoring, and the right to appeal. AI can contribute to improving educational quality and equity only when stakeholder responsibilities are clearly defined, decisions are explainable, and the entire process is documented with verifiable evidence.

5.2 Recommendations

5.2.1 For state management agencies

State management agencies should issue guidance specific to the education sector that defines high-risk use cases, standards for impact assessment, transparency requirements, and learners' rights to explanation and appeal. AI governance should be integrated with existing mechanisms for curriculum quality assurance and personal data protection.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

5.2.2 For higher education institutions

Higher education institutions should establish inventories of AI systems, issue internal regulations on AI use, and set up interdisciplinary AI review boards. Technology procurement contracts should specify data terms, audit rights, incident handling, and service termination. Every decision affecting admission, grading, academic progression, or graduation should have a designated person who bears final responsibility.

5.2.3 For instructors and learners

Instructors and learners need training in verifying AI outputs, citing AI use, protecting personal data, and recognizing the limitations of AI. Academic integrity policies should distinguish between permitted use, use that must be declared, and prohibited use. Learners should be provided with clear channels for requesting explanations and regrading.

5.2.4 For developers and providers

Developers and providers should test their systems for bias, disclose their limitations, ensure traceability, and allow human intervention. Systems should include a mechanism for suspension when risks exceed acceptable thresholds.

5.3 Limitations and future research

This study has several limitations. First, it is a conceptual and normative study based on a literature review and document analysis; the proposed framework has not yet been tested in practice. Second, Vietnam's legal framework for AI is new, and implementing regulations may alter some of the requirements discussed. Third, the analysis focuses on curriculum design and delivery and does not cover other uses of AI in higher education, such as research and administration. Future research should test the framework through case studies or pilot implementations at Vietnamese higher education institutions, survey the perceptions of lecturers, students, and administrators, and develop measurable indicators for each step of the framework.

REFERENCES

1) Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115–129. https://doi.org/10.1093/applin/14.2.115
2) Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9
3) Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027
4) European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689.
5) Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1
6) Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1
7) ISO/IEC. (2023). *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*. International Organization for Standardization.
8) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
9) Ministry of Education and Training. (2021). *Circular No. 17/2021/TT-BGDĐT on training program standards and the development, appraisal, and promulgation of training programs at all levels of higher education*.
10) Ministry of Science and Technology. (2026). *Circular No. 05/2026/TT-BKHCN promulgating the National Artificial Intelligence Ethics Framework*.
11) Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4
12) National Assembly. (2025a). *Law on Personal Data Protection No. 91/2025/QH15*.
13) National Assembly. (2025b). *Law on Artificial Intelligence No. 134/2025/QH15*.
14) National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1
15) OECD. (2024). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD Publishing.
16) Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM. https://doi.org/10.1145/3351095.3372873
17) Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366
18) Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
19) UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. UNESCO.
20) UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
21) Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/