Available on: https://jsshrs.ep-journals.org/ E-ISSN: 3051-3561, P-ISSN: 3051-3553

Journal of Social Science and Human Research Studies

Volume : 02 Issue : 11 November-2026

Responsible AI Governance in Higher Education Curriculum Design and Evaluation: An Integrated Model for Vietnam

Ha Thi Thu Thuy
Faculty of Economics, Hanoi Industrial and Trade University

Published Date: November 26, 2026 **Article DOI: 10.65150/EP-jsshrs/V2E11/2026-33**

ABSTRACT: As artificial intelligence, especially generative AI, becomes increasingly embedded in higher education, the key issue is no longer merely how AI can be used as a technical tool, but how it can be governed responsibly across the entire curriculum life cycle. This article employs an integrative literature review combined with policy analysis to clarify the theoretical foundations of responsible AI, educational data governance, outcome-based evaluation, and quality assurance. The study proposes a responsible AI governance model consisting of five components: educational purposes, data governance, AI model transparency, human oversight, and quality improvement. It also develops a risk-control matrix for AI use in curriculum design, implementation, and evaluation. Based on these findings, the article offers recommendations for Vietnamese higher education institutions seeking to adopt AI in a transparent, accountable, and quality-oriented manner.

KEYWORDS: responsible AI, curriculum, higher education, data governance, quality assurance.

1. INTRODUCTION

Artificial intelligence (AI) is reshaping the ways universities design, implement, and evaluate their curricula. Whereas earlier applications mainly involved adaptive learning, learning-outcome prediction, or automated grading, generative AI can now support the formulation of learning outcomes, course analysis, feedback synthesis, and recommendations for curriculum improvement. Under quality-assurance frameworks such as AUN-QA and ABET, curricula must demonstrate alignment among educational objectives, learning outcomes, course structures, pedagogical approaches, assessment, and continuous improvement (ABET, 2024; ASEAN University Network, 2020). As AI enters this cycle, evidence of quality increasingly includes learning and interaction data, stakeholder feedback, and skills forecasts alongside conventional programme documents and assessment results. Educational data governance and AI governance have therefore become important dimensions of curriculum management. In Vietnam, educational digital transformation and the national AI strategy have provided a policy foundation for the adoption of AI in higher education (Prime Minister of Vietnam, 2021, 2022). Implementation, however, cannot be driven by technological innovation alone. It must also comply with the legal framework, particularly provisions governing personal data protection, the purposes of data processing, data subjects' consent, and the protection of learners' information (Government of Vietnam, 2023).

Against this background, the article addresses three research questions: (1) What theoretical requirements should inform AI governance in higher education curriculum design and evaluation? (2) What are the major risks associated with AI use throughout the curriculum life cycle? (3) What responsible AI governance model could be appropriate for Vietnamese higher education? The study contributes by shifting the focus from simply 'using AI' to 'governing AI responsibly', thereby offering a complementary perspective on digital transformation and curriculum quality assurance. A curriculum-level perspective is essential because the consequences of AI-assisted decisions may accumulate across separate academic processes. A recommendation made during programme design can influence the courses offered, the evidence collected during delivery, and the criteria later applied in assessment. Accordingly, responsibility must be assigned across the full decision chain rather than delegated to a single lecturer or software provider. The analysis that follows treats AI outputs as provisional inputs to deliberation and treats approved programme documents, independently checked evidence, and accountable human decisions as the authoritative records. This distinction is especially relevant where institutions seek faster reporting while maintaining the evidential requirements of external quality assurance.

2. RESEARCH CONTENT

2.1. Research methodology

This study employs an integrative literature review combined with policy analysis. This approach is appropriate because it connects multiple bodies of literature, including AI in education, responsible AI, learning analytics, educational data governance, outcome-based education, and curriculum quality assurance. Unlike a narrowly focused systematic review, an integrative review enables the development of a conceptual model informed by diverse academic and policy sources.

Three groups of sources were selected. The first comprises academic studies on AI in higher education, generative AI, learning analytics, and digital-learning ethics. The second covers international policy frameworks and guidelines on responsible AI, notably those issued by UNESCO,

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

the OECD, and the European Commission. The third includes programme-accreditation and quality-assurance materials, particularly AUN-QA and ABET, together with Vietnamese policies on digital transformation in education, the national AI strategy, and personal data protection.

Sources were selected according to their direct relevance to educational AI or AI governance and their academic or policy value. Priority was given to works published from 2019 onward concerning generative AI and AI policy, while foundational studies on learning analytics and data ethics were retained. The material was analysed under four themes: (i) responsible AI principles; (ii) educational data and learners' rights; (iii) learning outcomes and programme accreditation; and (iv) oversight, evidence, and continuous improvement.

The analytical procedure consisted of four connected activities. First, each source was examined for the governance problem it addressed and the type of claim it supported, distinguishing normative principles, institutional recommendations, and empirical findings. Second, the extracted ideas were grouped by curriculum stage and by the party responsible for consequential decisions. Third, recurrent risks were matched with feasible preventive, detective, and corrective controls. Finally, the results were synthesized into the five-component model and the risk-control matrix. The analysis was interpretive: no pooled effect sizes, institutional prevalence rates, or comparative performance statistics were calculated. Consequently, the proposed relationships represent a reasoned conceptual synthesis rather than tested causal associations.

This method offers breadth but has limitations. The sources were purposefully assembled and were not selected through a reproducible systematic-review protocol; therefore, the paper does not claim exhaustive literature coverage or a formal risk-of-bias assessment. Policy guidance and accreditation criteria establish expectations but do not themselves demonstrate that a particular AI application improves educational outcomes. Likewise, evidence from one jurisdiction cannot be transferred to a Vietnamese university without considering its regulatory obligations, academic structures, data systems, and staff capacity. These distinctions guide the interpretation of the proposed model and prevent normative recommendations from being presented as observed implementation results.

2.2. Theoretical foundations of responsible AI governance in curricula

Responsible AI requires AI systems not only to perform accurately but also to conform to social values, human rights, fairness, transparency, and accountability. As Dignum (2019) argues, responsibility does not reside in an algorithm alone; it extends to the entire ecosystem of design, deployment, and use. This is particularly important in education because AI-assisted decisions can directly affect learning opportunities, assessment outcomes, and students' career pathways.

The European Commission High-Level Expert Group on Artificial Intelligence (2019) identifies seven key requirements for trustworthy AI: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination, and fairness; societal and environmental well-being; and accountability. These requirements support assessing educational AI not only for effectiveness but also for legality, ethics, and reliability. UNESCO (2023) likewise advocates a human-centred approach to generative AI that serves learning objectives and supports, rather than replaces, professional judgement. Its 2024 competency frameworks for teachers and students further emphasize AI ethics, technical understanding, human-centred thinking, and pedagogical competence.

From a data perspective, learning analytics can facilitate early warnings, personalization, and curriculum improvement. Nevertheless, transparency, consent, purpose limitation, and learner protection remain essential (Slade & Prinsloo, 2013). Accordingly, AI governance must be integrated with educational data governance. Where AI generates, analyses, or interprets quality-assurance evidence, AUN-QA and ABET requirements should also be extended to the quality of the underlying data and AI models.

For curriculum governance, the principles in Table 1 become operational only when translated into reviewable institutional practices. For example, human oversight requires a defined approval authority, a procedure for escalating disputed outputs, and a record of the version ultimately accepted. Transparency requires more than announcing that AI was used: it should identify the purpose of use, relevant information sources, the extent of automated processing, and the limits of the resulting recommendation. Fairness calls for examining whether data coverage and model performance differ across student groups, course formats, and modes of study. Such reviews must be proportionate to the decision's consequences and should avoid unnecessary collection of sensitive personal information.

Table 1. Responsible AI principles and their implications for curriculum governance

Principle	Implications for curricula	Required evidence
Human-centred approach	AI supports academic decisions without replacing programme committees, lecturers, or learners.	Approver-role rules, committee minutes, expert-intervention logs.
Transparency and explainability	Users should know when AI is used, which data are processed, and the limitations of its outputs.	Tool descriptions, processing purposes, model limitations, output-verification records.
Fairness and non-discrimination	Learning outcomes, course recommendations, and assessments must not disadvantage vulnerable learner groups.	Data-bias tests and impact reviews for different learner groups.
Privacy and data governance	Learning data must be collected for specified purposes, with access controls, security, and retention limits.	Data policies, data classification, consent mechanisms, access logs.
Accountability	Every major curriculum decision must have an accountable person and an auditable record.	Responsibility maps, appeals procedures, curriculum version histories, improvement reports.

Source: Author's synthesis of European Commission HLEG AI (2019), ASEAN University Network (2020), and ABET (2024).

2.3. Risks of AI across the curriculum life cycle

The curriculum life cycle typically begins with assessing social and labour-market needs, defining objectives and learning outcomes, designing course structures, implementing teaching and learning, evaluating outcome attainment, and periodically improving the programme. AI may contribute to each stage, although the nature and level of risk vary across stages.

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

During needs analysis, AI can process job advertisements, labour-market reports, and alumni feedback. The principal risks are non-representative data and biases relating to occupation, region, language, or type of enterprise. Mechanical reliance on such analyses may cause learning outcomes to follow short-term labour-market demand at the expense of academic foundations, civic competence, and lifelong learning.

When learning outcomes and course structures are designed, generative AI can help draft competency statements, map outcomes to courses, detect content overlap, and suggest resources. A key risk is 'outcome drift': AI-generated statements may sound plausible but lack disciplinary justification, fail to align with institutional missions, or prove difficult to measure. Programme committees must therefore validate each learning outcome against appropriate cognitive taxonomies, competency levels, and assessability.

During implementation, AI systems can personalize learning pathways, recommend resources, and help lecturers provide rapid feedback. Risks include rigid learner profiling, excessive surveillance, and restricted opportunities following inaccurate predictions. An AI-generated academic warning should be treated as a signal for support, not a final judgement of a learner's ability.

During assessment and improvement, AI can automate grading, analyse student work, synthesize feedback, and produce quality reports. This stage is particularly sensitive because assessment outcomes may affect grades, learning-outcome recognition, accreditation, and a programme's reputation. Consequently, AI assessment systems require accuracy and consistency criteria, explainability, and accessible appeals procedures.

A risk-based governance approach should therefore distinguish the severity of potential harm from the likelihood of an error and from the institution's ability to detect and correct it. Drafting a non-binding summary of anonymous course feedback can generally be reviewed before use, whereas using an opaque model to flag individual students for academic intervention may have immediate consequences for their opportunities and treatment. Controls should escalate accordingly, from source verification and routine editorial review to documented validation, restricted access, double-checking of consequential judgements, and an effective avenue for challenge. The proposed matrix is intended to guide this proportional allocation of oversight; it is not a validated numerical risk-scoring instrument.

Table 2. AI risk-control matrix across the curriculum life cycle

Stage	Potential AI use	Main risks	Controls	Quality evidence
Needs analysis	Analyse job advertisements, alumni feedback, and labour-market data.	Biased data, short-term focus, omission of foundational competencies.	Triangulate sources; expert-panel review; disclose data limitations.	Data-source reports, consultation minutes, validated skills lists.
Learning outcomes	Suggest outcomes; map competency frameworks and accreditation criteria.	Generic or unmeasurable outcomes; reliance on AI-generated wording.	Taxonomy review, measurability checks, programme-committee approval.	Outcome maps, PLO–CLO matrix, revision records.
Course design	Identify overlap, suggest resources, simulate course structures.	Unverified content, source bias, theory–practice imbalance.	Validate resources and citations; test with lecturers and learners.	Course outlines, reading lists, pilot feedback.
Teaching implementation	Adaptive learning, chatbots, virtual tutors, early alerts.	Over-surveillance, learner misclassification, less human interaction.	Disclose AI use, enable feedback, require lecturer confirmation for major interventions.	System logs, AI policies, learner-support reports.
Evaluation and improvement	Automated grading, outcomes analysis, accreditation-report synthesis.	Assessment errors, low explainability, disputed accountability.	Periodic model validation, sample double-marking, appeals, human approval.	Reliability reports, validation samples, programme-improvement records.

Source: Author's synthesis of Slade and Prinsloo (2013), UNESCO (2023, 2024), AUN-QA (2020), and ABET (2024).

2.4. Proposed integrated responsible AI governance model

Drawing on the preceding analysis, this study proposes an integrated model for responsible AI governance in higher education curriculum design and evaluation. Rather than viewing AI as an isolated tool, the model positions it within a curriculum-governance ecosystem comprising educational objectives, data, algorithmic models, human actors, and quality-assurance mechanisms. The five interrelated components form a continuous cycle.

The first component comprises educational purposes and learning outcomes. All AI applications must be anchored in the institution's mission, educational philosophy, learner needs, and agreed learning outcomes. AI is valuable when it improves the achievement of educational objectives, not simply when it produces more documents, data, or reports. Before deploying AI, institutions should therefore specify which curriculum-quality problem the technology is intended to address.

The second component is data governance, which underpins all AI applications. Data used in curriculum design and evaluation may include learning records, grades, learner feedback, graduate employment information, employer feedback, learning management system (LMS) records, and labour-market data. Each data category should be classified by sensitivity, processing purpose, and access rights. Personal data must be handled in compliance with applicable laws concerning lawful purpose, consent, retention, and security.

The third component comprises transparent and verifiable AI models. An AI tool should enter academic workflows only when the institution has a reasonable understanding of its functions, input data, limitations, potential biases, and conditions of use. Generative AI outputs

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

must be checked, rather than cited as unverified academic sources. AI-assisted assessment requires reliability testing, comparison with lecturers' judgements, and correction mechanisms when discrepancies are identified.

The fourth component is human oversight and accountability. AI must not be the ultimate decision-maker in curriculum governance. Programme committees, faculties, lecturers, examination units, and quality-assurance offices should have clearly assigned responsibilities at each stage. Decisions concerning learning-outcome approval, course revision, outcome attainment, or curriculum modification must be reviewed by accountable personnel and documented for traceability.

The fifth component is accreditation, feedback, and continuous quality improvement. AI can help shift quality assurance from periodic assessment towards more continuous monitoring. However, data and analytical results become valid quality evidence only when embedded in a criteria-based self-assessment process, informed by stakeholder feedback and linked to concrete improvement decisions. The model thus requires every AI application to be associated with a quality indicator and an explicit improvement mechanism.

The model is best interpreted as a system of necessary governance functions rather than five sequential technology-installation steps. Educational objectives define legitimate purposes; data governance determines what information may be processed; model assurance tests the reliability and boundaries of AI-supported analysis; human oversight authorizes action; and quality assurance evaluates whether the action achieved its intended educational purpose. Feedback from implementation should then trigger changes to source data, instructional practices, model configuration, or approval rules. This closed loop helps institutions separate an apparent improvement in reporting speed from a substantiated improvement in curriculum quality. It also makes it possible to discontinue a tool whose risks outweigh its educational value.

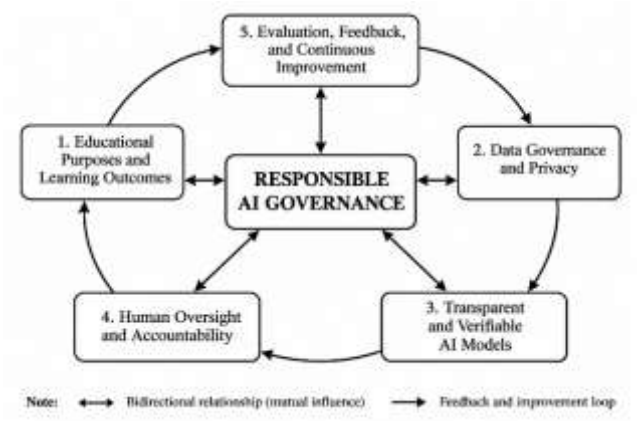

Figure 1. Responsible AI governance model throughout the curriculum life cycle
Source: Author's proposed model based on Dignum (2019), European Commission HLEG AI (2019), UNESCO (2023, 2024), AUN-QA, and ABET.

In practical terms, a proposed AI-assisted revision of programme learning outcomes can illustrate this interaction. The programme team should first specify the competency gap supported by employer, alumni, and academic evidence. An AI tool may then cluster the evidence and draft alternative outcome statements, but staff should verify the original sources, examine missing perspectives, and assess each draft for disciplinary relevance and measurability. The programme committee should retain the right to reject or revise every suggestion. Subsequent course mapping and assessment evidence should be reviewed against the approved outcomes, with a documented trail linking the initial problem, the recommendation, the human decision, and later evaluation. This example demonstrates how the framework may be applied; it is not a report of a pilot conducted by the present study.

2.5. Implementation recommendations for Vietnamese higher education institutions

First, universities should issue institutional regulations on AI use in curriculum design, implementation, and evaluation. These should go beyond simply permitting or prohibiting generative AI for learning. They should define permitted uses at programme, course, and assessment levels; assign institutional responsibilities; specify traceability and output-verification requirements; and establish procedures for dealing with errors and appeals.

Second, institutions should build interoperable yet controlled educational data-governance systems. Programme documents, courses, learning outcomes, student achievement, surveys, and stakeholder feedback should be standardized for analysis. Standardization does not justify collecting the maximum amount of information. Only data necessary for legitimate educational purposes should be collected, with sensitive information minimized, access controlled, and learner privacy protected.

Third, AI competence should be strengthened among lecturers, programme administrators, and quality-assurance staff. Many risks arise not only from the technology itself but also from misunderstanding or misusing it. Staff should be able to formulate appropriate prompts, verify outputs, identify bias, safeguard learner information, and explain how AI is used within a course or programme. UNESCO's 2024 competency frameworks can inform the design of professional development activities.

Fourth, AI applications should be piloted on a limited scale before broader implementation. Institutions may begin with lower-risk tasks, such as analysing course feedback, identifying overlap in syllabi, suggesting learning resources, or drafting self-assessment reports. Higher-risk applications-including automated grading, dropout-risk prediction, and proposed learning-outcome revisions-require pilots, independent evaluation, and human approval.

Fifth, AI should be linked to quality accreditation rather than treated as a technological trend. Each AI tool used in curriculum management should address three questions: Which quality criterion does it improve? What evidence demonstrates the improvement? Who is

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

accountable if its output is inaccurate? Answering these questions allows AI to become part of quality governance rather than merely a tool for content generation or data processing.

A feasible implementation sequence begins with governance readiness rather than procurement. During the preparation phase, a university can inventory existing programme data, identify high-consequence decisions, allocate roles, and define acceptable uses. In a bounded pilot phase, one programme or quality-assurance unit can test a low-risk use case with approved input sources, a pre-established review protocol, and clear criteria for discontinuation. During evaluation, reviewers can compare time expenditure, error types, traceability, stakeholder satisfaction, and the quality of the final approved output against the previous workflow. Expansion should follow only when the institution can show that benefits are meaningful, resource demands are sustainable, and identified risks remain controllable.

Institutional evaluation should avoid assuming that a higher quantity of AI-generated documents represents better quality assurance. Potential indicators include the proportion of AI-assisted statements with verifiable supporting sources, the frequency of material errors discovered during human review, completion of required approval records, the time needed to correct disputed outputs, and the proportion of decisions for which affected individuals have a clear review channel. For high-stakes applications, data representativeness and differences in error rates across relevant student groups require additional attention. These are proposed indicators for future piloting, not values measured in this paper.

3. CONCLUSION AND RECOMMENDATIONS

3.1. Conclusions

AI is increasingly important in higher education curriculum design and evaluation. Its value, however, lies not only in automation or data analytics, but also in how it is governed to serve educational purposes, safeguard learner rights, and enhance programme quality. This article has clarified the foundations of responsible AI, identified risks across the curriculum life cycle, and proposed an integrated governance model for Vietnam. The model consists of five components: educational purposes and learning outcomes; data governance; transparent and verifiable AI models; human oversight and accountability; and accreditation, feedback, and continuous improvement. Together, these components encourage a shift from 'using AI to work faster' to 'governing AI to work more appropriately, transparently, and responsibly'. For Vietnamese universities, this is an important way to capture AI's benefits while reducing data, ethical, quality, and legal risks. Future implementation should combine institutional policies, staff digital competence, data architectures, model-validation procedures, and learner participation. Only when these conditions are met can AI substantively support curriculum innovation and higher education quality improvement.

The study's principal contribution is therefore analytical and organizational: it links responsible AI principles to concrete curriculum decisions, identifies where failures might affect learners and programme quality, and specifies the documentary evidence needed for accountable oversight. Its principal limitation is the absence of empirical testing in Vietnamese institutions. Future research should use multi-programme pilot studies with clearly documented baseline practices, comparable decision tasks, independent output-quality assessments, stakeholder interviews, and analysis of implementation costs. Such designs could assess whether the integrated model improves traceability and decision quality, and under what institutional conditions it remains feasible. Until such evidence is available, the proposed framework should be treated as a structured guide for experimentation rather than a proven intervention.

3.2. Recommendations

The implementation priorities arising from the analysis are to establish institution-wide AI governance policies, standardize educational data under appropriate privacy safeguards, develop staff competence in responsible AI use, pilot high-risk applications under independent evaluation and human approval, and link every AI application to measurable quality-assurance criteria. These recommendations correspond to the five implementation measures detailed in Section 2.5.

For implementation, universities may designate a cross-functional review group bringing together academic programme leaders, teaching staff, quality-assurance officers, information-security personnel, and student representatives when appropriate. Its initial mandate should be limited and explicit: approving pilot use cases, documenting permissible data, establishing validation responsibilities, defining incident-response procedures, and reporting lessons from the pilot to the competent academic body. This arrangement avoids placing full responsibility for sensitive curriculum-related decisions on individual software users while preserving the authority of established academic committees.

REFERENCES

1) ABET. (2024). Criteria for accrediting engineering programs, 2025–2026. https://www.abet.org/wp-content/uploads/2024/11/2025-2026_EAC_Criteria.pdf
2) ASEAN University Network. (2020). Guide to AUN-QA assessment at programme level version 4.0. https://www.aunsec.org/application/files/9117/7942/9691/Guide_to_AUN-QA_Assessment_at_Programme_Level_Version_4.0.pdf
3) Government of Vietnam. (2023). Decree No. 13/2023/ND-CP dated 17 April 2023 on personal data protection. https://vanban.chinhphu.vn/?docid=207759&pageid=27160
4) Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, Article 22. https://doi.org/10.1186/s41239-023-00392-8
5) Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer. https://doi.org/10.1007/978-3-030-30371-6
6) European Commission, High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
7) Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf
8) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

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

Corresponding Author: Ha Thi Thu Thuy

9) Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson. https://edu.google.com/pdfs/Intelligence-Unleashed-Publication.pdf
10) Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

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

Corresponding Author: Ha Thi Thu Thuy