Artificial Intelligence, Assessment Integrity, and the Future of Curriculum and Teaching

Governing Educational Innovation Without Compromising Academic Standards

Sixbert SANGWA
ssangwa@openchristian.education
Open Christian Press, a publishing imprint of Open Christian University and Center for Faith and Work
Chapter DOI: https://doi.org/10.65655/ocp.m29.c103
Book DOI: https://doi.org/10.65655/ocp.m29

Abstract

Artificial intelligence (AI) changes the conditions under which universities generate knowledge, organise teaching, evaluate individual achievement and justify academic awards. Its significance is not exhausted by the efficiency of automated tools or the detection of misconduct: AI unsettles the relationship among intended competence, performed intellectual work, submitted evidence and accountable educational judgement. This chapter develops a Rwanda-anchored, critically evidence-based account of responsible AI integration across curriculum, pedagogy, assessment, institutional management and external review. A purposive documentary and conceptual synthesis combines foundational scholarship, recent experimental research syntheses on learning outcomes and assessment validity with UNESCO guidance, the National AI Policy, Rwanda's data-protection law, Higher Education Council instruments and international risk-management benchmarks. It differentiates generative, predictive, adaptive and increasingly agentic applications; examines disciplinary and role-specific AI literacy; and analyses authorship, academic integrity, accessibility, bias, privacy, procurement, academic labour and technological dependence. The chapter's distinctive contribution is an evidence-traceable governance model connecting educational purpose, legitimate authority, bounded AI use, independently interpretable assessment evidence, learner protection and reviewable decisions. Practical instruments include an institutional AI governance and academic assessment integrity framework, task-level disclosure template, assessment-redesign matrix, risk-based tool appraisal and worked hypothetical evaluation. These are original professional proposals, not regulator-approved or independently validated instruments. The chapter concludes that trustworthy AI-enabled higher education requires greater, rather than diminished, human responsibility for what is taught, learned, assessed and certified.

Keywords: artificial intelligence in higher education; assessment validity; academic integrity; AI literacy; Rwanda; institutional AI governance; evidence-traceable quality assurance

6.1 Introduction: AI as an educational, epistemological, institutional, and governance challenge

A student submits a persuasive policy analysis with accurate formatting, fluent transitions and citations that appear scholarly. A lecturer has never seen the student reason through the underlying evidence, and several references are nonexistent. Elsewhere, a student employs the same family of technologies to challenge assumptions, debug code and rehearse a difficult explanation, then independently defends the resulting work. Treating both cases as equivalent instances of “AI use” obscures the fundamental educational distinction. The first may provide unreliable evidence of learning and knowledge; the second may represent intellectually demanding, properly acknowledged assistance. The central question is therefore not whether a university should welcome or reject a technology in the abstract, but whether its practices preserve justified claims about knowledge, authorship, competence and educational responsibility (Kane, 2013; Miao & Holmes, 2023).

The first five chapters of this volume established educational purpose as the reference point for quality, regulatory authority as a necessary constraint, institutional capability as the basis for sustainable delivery, constructive alignment as the architecture of curriculum, and pedagogical design as the condition of credible online and blended learning. AI adds a problem that cuts across these domains. It can participate in the production of teaching materials, explanations, student submissions, feedback, grades and evaluative reports. Once the same tool influences both the preparation of evidence and the judgement made from that evidence, the university must examine not merely accuracy but the independence, accountability and limits of the entire evidentiary chain (Sangwa, 2026; Zawacki-Richter et al., 2019).

The problem has four dimensions. Educationally, AI may change which competencies matter and which activities cultivate them. Epistemologically, plausible machine-generated language and synthetic evidence can be mistaken for warranted knowledge, especially where verification is weak. Institutionally, software contracts, staff expertise, student access and information flows alter who exercises practical control over teaching and assessment. Normatively, a university has to reconcile academic freedom, learner rights, qualified human judgement, domestic law, official HEC guidance and appropriate international benchmarks. These dimensions interact but cannot be collapsed into a general instruction to “use AI responsibly.”

Rwanda provides a particularly instructive policy setting. The Ministry of ICT and Innovation's National AI Policy articulates a national development direction, while the Higher Education Council's November 2025 Guidelines and Assessment Tools for the Virtual Learning and the Use of Artificial Intelligence in Rwanda's Higher Learning Institutions address AI-specific curriculum, staff readiness, student literacy, assessment rules, disclosure, institutional oversight and an evaluation instrument. HEC's March 2026 distance-learning guidance adds modality-specific educational expectations, and Presidential Order No. 044/01 of 24 July 2026 sets out the contemporary mandate of the Council. The HEC document presents its provisions as official guidelines and assessment standards and states that implementation follows official publication and dissemination, while also encouraging institutional alignment. Its specific applicability and relationship to superior legal instruments require source-sensitive interpretation; these sources must be distinguished by legal status, date, scope and competent authority; a professional framework in this chapter cannot amend any of them (HEC, 2025, 2026; Republic of Rwanda, 2026).

This chapter adopts a purposive, critical documentary synthesis rather than claiming a systematic review or field study. The source base comprises the author's doctoral analytical architecture, the completed preceding chapters, selected foundational and contemporary peer-reviewed scholarship, official Rwandan legal and regulatory material, UNESCO's human-centred AI guidance, and the US National Institute of Standards and Technology's voluntary risk-management guidance. Sources were selected for conceptual explanatory value, direct normative relevance and their capacity to test optimistic and restrictive positions. No new searches with a preregistered protocol, exhaustive screening counts, primary interviews, student data, field audits or effectiveness trials are claimed. References support individual propositions; the practical instruments and integrative model are the author's reasoned proposals, and all cases expressly identified as hypothetical are constructed illustrations.

An appropriate balance requires rejecting both technological inevitability and technological exceptionalism. A university need not deploy a particular commercial system merely because its competitors do so. Neither can it preserve educational standards by assuming that forbidding an embedded technology will make it absent. Good governance asks which educational functions warrant AI assistance, which require unaided individual performance, which risks cannot be accepted, and what evidence would allow an independent reviewer to decide whether the arrangements actually work. The answer must remain revisable as models, scholarship, costs and regulatory provisions change.

A further distinction concerns the status of the knowledge claims being made. A fluent answer is an observable linguistic performance, whereas a graduate's understanding is a capacity inferred from appropriately chosen evidence. A university cannot move directly from the former to the latter without examining authorship, task conditions, independent reasoning and the possibility of alternative explanations. The critical realist orientation developed earlier in the volume reinforces this difference among an observed artefact, the educational practices that generated it and the underlying capabilities that a programme claims to develop. AI makes the distinction harder to ignore, but the problem also applies to traditional ghostwriting, unacknowledged collaboration and examinations that reward memorised responses. The chapter therefore treats AI governance as part of ordinary academic quality governance, albeit with newly demanding technical and institutional manifestations (Kane, 2013; Sangwa, 2026).

6.2 Historical evolution and educational applications of artificial intelligence

AI in education did not begin with publicly accessible conversational systems. Earlier research explored symbolic reasoning, rule-based expert systems, programmed tutoring and models of how learners solve problems. Later systems used statistical learning to identify patterns in student answers, recommend exercises, classify submissions and forecast disengagement. Educational institutions also used algorithms in administration, information retrieval and institutional research long before generative interfaces made AI conspicuous to ordinary users. This history matters because different AI applications pose different questions about validity, bias and accountability; a plagiarism classifier, adaptive tutor and language model should not be evaluated as though they are interchangeable technologies (Zawacki-Richter et al., 2019).

The growth of machine learning shifted emphasis from explicit rules written for narrow tasks toward systems whose behaviour depends substantially on patterns inferred from data. Developments in natural-language processing and large-scale neural networks expanded the capacity to generate language, software, images and other artefacts. Foundation models and multimodal interfaces now make assistance available through ordinary learning platforms, office suites and search tools. Their usefulness is significant, but the appearance of fluent dialogue should not be confused with human understanding, reliably grounded reasoning or a guarantee of factual accuracy. A generated answer may be correct, incorrect, incomplete or misleading, often without signalling which condition obtains (Bender et al., 2021; Kasneci et al., 2023).

An important analytical distinction separates prediction, generation and action. A predictive system produces an estimate, such as the likelihood that a student may miss a progression milestone. A generative system produces an artefact, such as a worked example or draft paragraph. An agentic system can pursue a multi-step objective through tool use, retrieval, external applications or delegated actions. Each increases the potential distance between an initial human intention and the resulting institutional consequence. Prediction raises questions about statistical validity and the ethics of intervention; generation raises questions about truth, attribution and the assessment construct; action raises

questions about authorisation, security, limits and rollback. The risks accumulate when an agent can access records, communicate externally or modify academic systems.

Earlier systematic scholarship already identified an imbalance between technical innovation and educational involvement. Zawacki-Richter et al. (2019), reviewing the pre-generative-AI literature, found substantial attention to profiling and prediction, assessment and evaluation, adaptive systems and intelligent tutoring, while raising questions about the relative marginality of educators in some research. The result should not be extrapolated to a measured 2026 adoption rate. Rather, it supplies a historically situated warning: an education system can acquire increasingly powerful analytical tools without adequately resolving who frames the educational question or evaluates the learner's experience.

For curriculum teams, historical perspective prevents two recurrent errors. The first is assuming that novelty establishes educational merit; many supposed innovations revive longstanding ambitions for personalisation, automation or reduced teaching cost. The second is applying a single AI rule to dissimilar applications. A speech-to-text accommodation, a simulated clinical conversation, a source-retrieval system, an automated essay scorer and an autonomous admissions workflow differ in data requirements, pedagogical purposes, failure consequences and appropriate oversight. Each should be evaluated as a specific socio-technical use case rather than approved under a generic institutional declaration that the university “uses AI.”

The intellectual history also contains competing conceptions of what it means to individualise teaching. Rule-based tutoring sought to represent domain knowledge and learner error explicitly; data-driven recommendation increasingly estimates patterns from previous interactions; contemporary language-model assistance can simulate an explanatory conversation without possessing a validated representation of the learner's conceptual state. These approaches may coexist within the same university. The difference matters when a lecturer interprets a system's statement that a student has achieved mastery. A diagnostic label might derive from a relatively interpretable exercise bank, an opaque prediction model, or a conversational response that is only loosely connected to assessment standards. Before making decisions about progression, an institution should determine which kind of inference the technology actually supports and which corroborating human assessment is necessary. Historical innovation does not dissolve the need to identify the source of a learning claim.

6.3 Generative AI, agentic systems, adaptive learning, automated feedback, and intelligent tutoring

Generative systems may support brainstorming, explanation, translation, code drafting, formative questioning, scenario creation and accessibility, particularly when students can challenge and verify outputs. Their principal educational value is conditional: an explanation becomes useful when the learner interrogates it against disciplinary concepts and credible sources; feedback becomes useful when it accurately addresses the student's work and prompts subsequent revision. The technology can also invite cognitive substitution, whereby a plausible completed artefact replaces the very reasoning the activity was intended to develop. Because these outcomes can arise from the same interface, evaluations should focus on the learning task and evidence of student action rather than the marketing category of the tool (Kasneci et al., 2023; Miao & Holmes, 2023).

Adaptive learning systems typically adjust material, pace or suggested practice based on learner information. Such adjustment can help novices encounter suitably challenging exercises and receive additional practice. Yet a system's inferred “mastery” is only as defensible as the instrument, learning model and data on which it depends. Repeated correct answers to narrow questions may establish local procedural fluency without demonstrating transfer, disciplinary explanation or professional judgement. Institutions should ask how the system estimates progress, what it does for learners whose profiles depart from training data, and whether staff can see and correct inappropriate recommendations. Personalisation should not become a quiet mechanism for assigning students to different intellectual ceilings.

Intelligent tutoring can provide guided dialogue or stepwise practice that would otherwise be scarce. A well-designed use may require students to state a preliminary answer, expose assumptions, compare alternatives and reflect on errors before receiving further explanation. Conversely, an unstructured tutor that supplies complete solutions immediately can reduce productive struggle. Universities should therefore distinguish assistance that scaffolds thinking from assistance that bypasses it. The available research supports plausible opportunities and documented limitations, not a general conclusion that conversational tutoring consistently outperforms expert human teaching across disciplines and populations (Zawacki-Richter et al., 2019; Kasneci et al., 2023).

Automated feedback similarly requires disaggregation. Spelling correction, detection of missing submission elements, rubric-aligned comments and the evaluation of a complex legal or ethical argument have different evidentiary demands. Machine assistance can help lecturers produce an initial set of comments or identify patterns, but an apparent explanation may misunderstand a student's purpose, penalise unconventional but defensible thinking, or cite nonexistent evidence. The lecturer remains responsible for the fitness of feedback in relation to the stated criteria. Where AI materially contributes to high-stakes grading, the institution should require evidence of suitability, consistency, contestability and actual human review rather than merely naming a supervisor.

Agentic AI enlarges the stakes. A system instructed to “prepare the week's feedback” might retrieve assignments, rank students, draft individual responses and post them automatically. Such a sequence creates compound risks: incorrect access permissions, exposure of confidential records, biased ranking, misleading feedback and irreversible publication. Effective controls include limited permissions, approved destinations, separated approval steps, transaction logs, testing in nonproduction environments, prompt-injection resistance, and a named human who can halt or reverse actions. “Human in the loop” has little meaning when the operator cannot inspect what the agent did or lacks time and authority to intervene (Tabassi, 2023; Autio et al., 2024).

Institutional adoption should proceed from a use-case record specifying the intended educational benefit, required data, affected learners, failure modes, independent evidence of accuracy, accessibility, cost, residual risk, oversight and exit plan. Table 6.1 makes these distinctions operational. It is not a ranking of technologies. A relatively simple classifier may be high risk when it shapes admission or disciplinary outcomes; a sophisticated generative tutor may be lower risk when used voluntarily for reversible, supervised practice.

Table 6.1

Illustrative AI applications and the appropriate locus of educational control

Application	Potential contribution	Principal evidentiary question	Appropriate control
Predictive analytics	Identify learners who may need support	Is the prediction valid, fair and useful for this cohort?	Human review, consent/legal basis and support safeguards
Generative study aid	Provide explanations and examples	Can learners verify outputs and still demonstrate understanding?	Task instructions, source checking and accessible alternatives
Adaptive tutor	Provide differentiated practice	Does suggested practice support the intended mastery?	Educator oversight and performance checks
Automated feedback	Accelerate formative response	Are comments accurate, actionable and consistent with standards?	Sampling, staff moderation and correction route
Agentic workflow	Execute multi-step administrative tasks	Can every material action be authorised, inspected and reversed?	Least privilege, approval gates, logging and rollback

Note. Original analytical classification by the author, informed by Zawacki-Richter et al. (2019), Miao and Holmes (2023), and Tabassi (2023) and Autio et al. (2024). Controls are proposed, not official risk designations.

Contemporary intervention evidence warrants more explicit attention than an unqualified narrative of either improvement or intellectual decline. Deng et al. (2025) synthesised 69 experimental studies of ChatGPT-supported learning and reported favourable results for several measured outcomes, while cautioning that small samples, limited statistical power, outcome-selection practices and testing under conditions of continued AI access complicate interpretation. Chen and Cheung (2025), in a meta-analysis of 57 studies of generative AI in university learning, reported positive average results across several outcome categories but did not find a statistically significant metacognitive effect. Their findings are not interchangeable estimates of one universal educational benefit: the research questions, populations, intervention conditions, outcomes and pooling decisions differ. Nor do improvements on intervention-aligned tests necessarily demonstrate durable independent capability. These syntheses provide an evidence-based reason to pilot well-designed AI support, not a warrant to replace disciplinary instruction or assessment verification across the board.

A particularly informative, though not peer-reviewed, working paper by Contractor and Reyes (2026) reports a randomised undergraduate experiment using immediate and delayed unaided assessments. It distinguishes students who sought explanations from those who primarily automated text production, observing that the former tended to retain more of the subsequent unaided writing benefit. The experiment is limited to its specified tasks, participants and follow-up period, and the authors' proposed mechanisms should not be treated as universally established. Nonetheless, the design illustrates an important methodological standard: evaluation of an educational AI intervention should test learning after assistance is withdrawn when independent performance is part of the stated outcome. A university introducing AI tutoring should therefore monitor what students can explain or perform later, not merely whether the system helped them submit a stronger immediate product (Contractor & Reyes, 2026).

6.4 AI and curriculum transformation: Graduate competencies, disciplinary knowledge, and professional readiness

AI makes curriculum review more urgent but does not justify replacing an intellectually demanding degree with a collection of application tutorials. Professional tools change more quickly than foundational concepts, and graduates need the means to evaluate technologies they did not encounter during their studies. Curriculum therefore has to

differentiate enduring knowledge, general AI literacy, discipline-specific applications, professional judgement and capabilities that remain indispensable when technological services fail. A marketing student may use generative systems to draft options, but still needs to understand evidence quality, consumer behaviour, research design, competition and ethical representation. A health student may use decision support but must distinguish a suggestion from clinically accountable judgement. A historian can explore machine-transcribed archives yet must retain source criticism and interpretive responsibility (Biggs, 1996; Miao et al., 2024).

The appropriate unit of change is the graduate profile and programme learning outcomes, not the isolated addition of a fashionable module title. A programme should identify which graduates must understand AI as an object of study, which must use AI as a tool, which will design or govern AI systems, and which must recognise when AI is inappropriate. These roles overlap but require different levels of mathematics, computing, disciplinary knowledge and professional accountability. An introductory AI-literacy component can establish common foundations; repeated situated learning must then show how those foundations operate in research, teaching, law, business, engineering or community service. Without disciplinary application, a cross-cutting module risks producing generic vocabulary with little durable competence.

Curricular mapping should track at least five capabilities: understanding the purposes and limits of relevant AI systems; formulating meaningful problems and prompts; evaluating accuracy, provenance and uncertainty; managing privacy, bias, attribution and safety; and deciding when to reject, escalate or override an output. For programmes creating AI, additional mathematics, data engineering, validation, security and system-design outcomes are indispensable. Graduates should encounter counterexamples and real failure modes. Assessment of AI literacy cannot rest only on a student's ability to obtain an attractive response because high-quality prompting may coexist with poor verification or shallow conceptual understanding (Miao et al., 2024; Miao & Cukurova, 2024).

Curriculum redesign also raises a question of opportunity cost. Every new learning outcome consumes time, assessment capacity and sometimes specialised equipment. Academic boards should examine whether redundant content can be removed without weakening knowledge progression; whether AI can be integrated into existing cases; whether faculty can deliver the revised material; and whether any profession imposes distinct practice requirements. Curriculum committees should resist a practice of placing “AI” in many module descriptions while leaving no assessed evidence of independent professional judgement. Such symbolic alignment is particularly dangerous where the institution advertises AI-ready graduates before building corresponding teaching and validation capacity.

A useful three-stage developmental structure distinguishes *foundations*, *critical application* and *accountable independent performance*. Early learners explain how an output may be produced, identify uncertainty and apply simple data-protection rules. Intermediate learners compare AI and non-AI approaches, document verification and explain bias or failure. Graduating learners exercise contextual judgement in a consequential task, defend choices without depending upon the original system, and show that the final product meets disciplinary standards. Progression should be assessed across modules through an outcome map that also records the tasks students must complete without assistance. This preserves the meaning of the qualification rather than equating technological fluency with competence.

Rwanda's HEC 2025 guidance specifically calls for review of module outcomes, learning activities and resources and for a cross-cutting AI-literacy module reviewed at least annually, together with staff and student preparation. The National AI Policy provides developmental context but does not by itself prescribe a particular degree structure. Institutions should therefore verify the presently applicable HEC instrument, map its provisions to approved programmes, and then defend their discipline-specific curriculum choices through academic governance. The professional challenge is to move from apparent compliance to evidence that each graduate has acquired the knowledge and judgement described in the revised award (HEC, 2025; Ministry of ICT and Innovation, 2023).

Graduate preparedness also requires a distinction between professional competence with AI and dependence on a specific vendor's interface. Consider a hypothetical public-health programme. Graduates may need to inspect the provenance of epidemiological data, recognise selection bias, critically examine an automated risk summary and communicate uncertainty. They need not all develop foundation models. Conversely, a computer-science programme that promises AI engineering capability must go beyond prompt formulation to mathematics, model development, evaluation, software testing, security and reproducibility. Law, humanities, business and education programmes make different legitimate choices. The graduate profile should justify these choices against disciplinary standards and foreseeable practice, indicating which competences are foundational, where AI creates new professional responsibilities, and which essential capabilities must remain demonstrable without AI. Curriculum revision becomes rigorous when it names both what will be added and which existing learning will be protected.

Programme approval should also consider how the curriculum represents ethical conflict rather than simply endorsing a set of desirable behaviours. Students may encounter a system that saves staff time but has uneven error rates; a helpful translator whose use exposes confidential research data; or an automated recommendation inconsistent with a client's stated priorities. A demanding programme asks students to identify the competing obligations, evaluate alternative methods, justify an appropriate decision and explain residual uncertainty. Generic instruction to 'use AI ethically' cannot replace assessed opportunities to make these judgements. It is particularly important to distinguish the correct operation of a tool from professional accountability for its consequences (UNESCO, 2021; Miao et al., 2024).

6.5 AI literacy and competency development for students, lecturers, academic leaders, and reviewers

AI literacy is not a single digital skill and should not be measured solely by confidence or frequency of use. Early conceptual research differentiated understanding, use, evaluation and ethical engagement rather than equating literacy with software operation (Ng et al., 2021). Students must know what relevant tools can and cannot establish, practise independent verification, recognise ethically consequential data, and explain the intellectual contribution they make when assisted. UNESCO's 2024 competency framework for students identifies a human-centred mindset, ethics, AI techniques and applications, and AI-system design, with progressive stages of understanding, application and creation. Although designed as a curricular resource rather than a Rwanda-specific regulation, it provides a useful basis for differentiating elementary user awareness from genuinely critical and creative agency (Miao et al., 2024).

Lecturers require an additional pedagogical layer. They need to distinguish constructive assistance from substitution, redesign outcomes and assessment tasks, verify machine-generated examples, detect fabricated references, evaluate learner accessibility and understand disciplinary limits. UNESCO's teacher framework presents human-centred orientation, AI ethics, foundations and applications, AI pedagogy, and professional learning as related dimensions. It recognises that the lecturer is not simply a consumer of educational software but an accountable professional whose decisions shape student agency. Training in writing prompts is insufficient if lecturers cannot establish whether a proposed use improves or compromises learning (Miao & Cukurova, 2024).

Academic leaders, programme chairs and regulators face different competence requirements. A dean must understand governance authority, procurement risks, academic workload, appeal rights and the evidentiary limitations of automated reports. A registrar must protect award records and student information. A research leader must address confidential datasets, attribution and publication ethics. An HEC programme evaluator needs enough literacy to interrogate claims of AI integration, recognise false proxies for readiness and request appropriate demonstrations without pretending to be an expert in every model architecture. The central professional skill for reviewers is the ability to connect a specific educational claim to an applicable rule, a relevant evidence source, an adequate verification test and a bounded evaluative conclusion.

Professional development should therefore be role-based, iterative and assessed through authentic performance. Staff might be asked to revise an assignment brief with task-specific permissions, identify fabricated citations in a generated literature summary, assess an accessibility failure, or document why a proposed predictive system should not be procured. Participation certificates show that learning opportunities were offered; they do not establish competence. More persuasive evidence includes moderated artefacts, trained reviewers' performance in realistic scenarios, improvement in assessment instructions and verified changes to institutional procedures. Continuing updates are required because platforms, institutional policies and misuse patterns change, but training frequency alone is not a meaningful educational outcome.

Equitable development requires attention to language, disability, connectivity, cost and discipline. An employee with unreliable access should not be judged deficient because the institution has imposed a subscription-based tool without alternatives. Students should not need to expose sensitive personal circumstances to obtain an AI-enabled accommodation. Training materials should demonstrate tool use and non-use, include low-bandwidth routes and offer assessment methods that do not reward only those with premium subscriptions. In multilingual contexts, institutions should investigate whether examples and outputs reflect the language varieties, disciplinary sources and local practices of their learners rather than treating English-dominant results as neutral (UNESCO, 2021; Liang et al., 2023).

Evidence of literacy must ultimately be visible in work. A credible learner can explain why a response is misleading, identify an unsupported premise, find an authoritative contrary source and decide whether the tool ought to be used at all. A credible lecturer can defend the assessment inference after permitted AI assistance. A credible reviewer can distinguish HEC's stated requirements from professional recommendations and explain the limits of an inspection result. This conception of literacy connects technological skill to epistemic responsibility and prevents the institution from measuring success by adoption figures alone.

A practical way to assess competence is to examine comparable work under three conditions: without AI for indispensable foundational tasks, with bounded AI for supervised application, and with a deliberately flawed or uncertain AI output requiring independent evaluation. A postgraduate research student, for example, should identify an appropriate design without relying on generated methodological language, use an approved assistant to compare alternative analytical plans, and finally detect ungrounded causal claims in a synthetic interpretation. The evidence should be moderated against transparent criteria rather than scored according to the sophistication of the prompts alone. This approach avoids treating AI literacy as a ceremonial induction certificate and enables institutions to distinguish tool proficiency from conceptual mastery.

The same principle applies to staff and reviewer competence. Academic developers should use observed work, including a corrected assessment brief, an independently checked reference list and a documented response to an AI security or fairness failure, rather than infer capability from workshop attendance. A reviewer evaluating an institution's AI training programme should ask for a sample of staff-produced artefacts, the methods used to judge them and corrective

actions where misconceptions persisted. Limited sampling cannot prove the competence of every employee, but it is stronger evidence of enacted practice than a schedule or attendance record alone. Review sampling should be risk-sensitive and should protect confidential academic materials (Miao & Cukurova, 2024; Sangwa, 2026).

6.6 AI-assisted curriculum design, lesson planning, resource development, teaching, and academic research

AI can accelerate preliminary curriculum work by suggesting an inventory of knowledge domains, comparing phrasing of intended learning outcomes, drafting formative cases and checking whether a module description contains obvious gaps. The efficiency is real in the limited sense that text can be produced rapidly. Educational quality, however, depends on whether these suggestions survive disciplinary scrutiny, qualification-level evaluation, stakeholder evidence, workload testing and regulatory review. A generic system asked to create a bachelor's curriculum may reproduce fashionable module labels, inappropriate prerequisites or imaginary professional regulations. The academic team must therefore treat output as a proposal whose provenance and suitability remain to be established, not as a validated academic specification (Biggs, 1996; Miao & Holmes, 2023).

Responsible curriculum assistance begins with authoritative input and clear task limits. A committee may supply approved programme learning outcomes, the relevant national qualifications descriptors, an institution's academic regulations and a structured list of existing modules, provided that no protected information is exposed to an unauthorised service. The AI system can be asked to identify candidate inconsistencies or generate alternative sequencing scenarios. Staff then manually verify credit calculations, curricular dependencies, qualification demand and cited provisions against the originals. The final decision record should identify which suggestions were accepted, rejected or revised and why. Such a procedure preserves a defensible human academic rationale while using automation for preliminary comparison and drafting.

Lesson planning offers similar opportunities and hazards. A lecturer may use AI to produce multiple explanations of an abstract concept, generate a contrasting case, suggest accessible alternatives for a complex diagram or create questions for formative practice. These resources require verification for factual correctness, linguistic suitability, alignment with intended learning and accessibility. In a quantitative discipline, a generated worked solution should be solved independently; in law, every legal proposition should be checked against current authority; in theology or history, quotations and bibliographic claims should be located in genuine texts; in health education, scenarios must be clinically reviewed and de-identified before use. The appearance of a plausible teaching resource is not evidence that it can safely be released to learners.

The research environment raises particular concerns. AI may help construct search expressions, organise permissible notes, explain statistical code or suggest manuscript structure. It cannot be assumed to have searched an identified database merely because it reports studies or produces citations. References, quotations, datasets, calculations and analytical claims must be independently verified. A researcher may not invent search or screening records, fabricate an interview transcript, misrepresent machine-produced summaries as direct reading, or transfer confidential peer-review documents into a public service without permission. Research authors retain responsibility for methodological transparency, scholarly originality and publication ethics, whether or not AI materially assisted a stage of their work (UNESCO, 2021; Miao & Holmes, 2023).

There is also a danger of epistemic homogenisation. If academics increasingly rely on similar systems to nominate readings, generate examples and draft assessment prompts, curricula may converge around highly represented material while neglecting African scholarship, local languages, community knowledge and disciplinary dissent. This is not a claim that all AI systems systematically omit African sources. It is a risk requiring active testing and the deliberate inclusion of locally relevant material. Programme teams should compare suggested sources against scholarly databases, African repositories, specialised professional literature and relevant regulatory instruments. Intellectual diversity depends upon informed curation, not the volume of automatically generated resources.

Institutional quality assurance should examine how AI-assisted content is checked before release and after student feedback. A feasible workflow separates generation, academic verification, copyright and privacy review, accessibility testing, approval, publication and periodic correction. The verifier should be competent in the subject and able to identify plausible errors; the reviewer should retain a record of the material actually approved. For continuously generated student-facing explanations, a different approach may be needed: boundary prompts, curated knowledge bases, clearly identified uncertainty, reporting routes, transcript sampling where lawful, and escalation to a qualified human. Such controls should be proportionate to consequence rather than imposed identically on a low-stakes study quiz and high-stakes clinical instruction.

For curriculum committees, it is helpful to organise AI-assisted design as a controlled academic-production process. The initiating lecturer specifies the intended learning, audience, relevant authoritative resources, confidentiality classification and the exact function delegated to the tool. The generated candidate material is then checked by a competent subject specialist against original sources, tested for cognitive demand and accessibility, and examined for assessment conflicts. The module coordinator records the adopted version and a date for review; where the material is

used in summative assessment, appropriate moderation and security controls apply. Each stage needs a named owner because distributed use of a tool can otherwise make a factual or pedagogical defect nobody's responsibility. The evidentiary record should be concise enough to maintain in ordinary departmental work yet sufficient to reconstruct consequential decisions.

AI-supported research also raises a conflict between the convenience of synthetic material and the integrity of evidentiary provenance. A generated bibliography is not the result of a database search, an invented dataset is not a field observation, and a confident summary does not demonstrate that the cited source was examined. Similarly, software-generated coding suggestions or statistical interpretations must not be presented as independently verified analysis until the researcher has checked inputs, assumptions and results. When confidential manuscripts, institutional records or student research are processed, privacy and contractual authorisation are prior conditions rather than formalities to address after publication. Editors, supervisors and examination boards should expect proportionate disclosure of material assistance while retaining ordinary criteria for originality, validity and research ethics (Miao & Holmes, 2023; UNESCO, 2021).

6.7 Assessment validity, authentic assessment, higher-order reasoning, and evidence of individual learning

The fundamental question of assessment is inferential: what may the university legitimately conclude about a learner from the performance it observes? An essay grade might support a claim about source-based reasoning if the student conducted that reasoning, the task elicited it and the scoring criteria reflected it. When an AI system supplies the literature synthesis, structure, argument or final answer, the identical grade may support a different and narrower claim. AI does not automatically invalidate an assessment; it changes the assumptions that connect a submitted artefact to an interpretation of individual competence. Kane's (2013) argument-based account of validity is particularly useful because it requires scrutiny of the evidence and inferential steps supporting the intended interpretation and use of scores.

Constructive alignment remains necessary but is insufficient when the learner's contribution is unobserved. A module may align the stated outcome “critically evaluate competing explanations” with an essay called “Critical evaluation,” yet the assignment may be completed largely through unexamined machine synthesis. Assessment designers must establish what quality of individual reasoning needs to be demonstrated, whether permitted AI assistance threatens that evidence, and what complementary task would make it visible. The solution may be a short oral defence, a supervised analysis of unfamiliar material, a staged decision log, a live debugging exercise or triangulation of several performances. The choice follows the construct, not a universal rule that every student must sit an additional examination (Biggs, 1996; Kane, 2013).

The distinction between product and process is consequential. A polished final artefact may be valuable in professional practice, especially where responsible use of tools forms part of real work. Yet it can conceal the origin of judgements, the handling of uncertainty and the student's ability to transfer knowledge to a new situation. Process evidence should make relevant thinking inspectable through research notes, interim decisions, revised assumptions, selective draft histories, code commits, laboratory notebooks or reflective explanations. Such evidence is not infallible: logs can be fabricated, oral defences can privilege confident speakers and controlled examinations can create disability or language disadvantages. A valid design uses several appropriate sources without equating procedural inconvenience with academic rigour.

Authenticity likewise requires conceptual precision. A “real-world” task is not inherently valid because it resembles a workplace; the task must represent the competence being claimed, and the scoring must distinguish essential performance from irrelevant circumstances. A simulated procurement decision can test reasoning about supplier evidence, risk and public accountability if those dimensions are carefully represented. An unsupervised group project may be authentic to teamwork but provide weak evidence of each member's individual statistical competence. Conversely, a conventional written examination can be appropriate when independent explanation or foundational calculation is indispensable. Authenticity is an alignment claim about meaningful performance, not a decorative attribute attached to case studies.

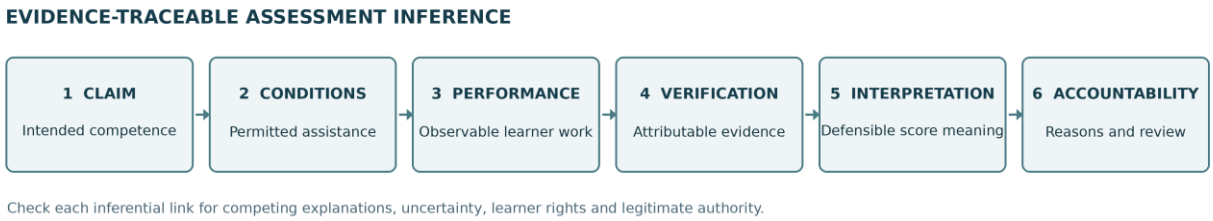
Assessment in an AI-rich environment must differentiate tasks where AI is prohibited because unaided competence is the construct, tasks where bounded assistance is permitted, and tasks where critical use of AI is itself assessed. These categories require compatible marking criteria. If AI is encouraged to generate preliminary options, students should be evaluated on verification, selection, explanation and accountability, not merely on the fluency of the resulting output. If the assessment concerns foundational statistical reasoning, the student must demonstrate the relevant operations or interpretation independently. This is particularly important in professional and safety-critical programmes where a correct machine-assisted answer does not necessarily show that a graduate can recognise a dangerous error.

Programme-level assurance should test the distribution of assessment evidence across an entire award. A single high-stakes capstone cannot automatically compensate for three years of unclear authorship, while a large number of low-level quizzes may fail to demonstrate graduate-level synthesis. Curriculum committees need an assessment blueprint

indicating which outcomes are assessed, at which developmental stage, with what AI conditions, how individual learning is authenticated, and where moderation occurs. The record should distinguish intended design from actual assessed work and the academic decisions made from it. The essential result is credible graduate certification rather than maximal surveillance or the accumulation of detector scores.

Figure 6.1

From student performance to a defensible qualification claim



Note. Original conceptual diagram. A submitted artefact is evidence only through a chain of explicit claims, authorised assistance conditions, verification, scoring, interpretation and review; it is not direct proof of independent achievement.

An assessment argument can be decomposed into several propositions that require different controls. Scoring asks whether the observed response has been evaluated consistently against the intended criteria. Generalisation asks whether performance would remain defensible across appropriate tasks or examiners. Extrapolation asks whether the result supports a claim about capability outside the immediate task. Consequential use asks whether a grade, progression decision or professional credential is justified in light of that evidence and its limitations. AI can disturb any of these connections: an automated marker may mis-score an original answer; an assisted essay may inflate apparent individual performance; a polished simulation may fail to predict authentic practice; or a detector-driven misconduct action may generate serious harm from weak evidence. Argument-based validation helps institutions select targeted safeguards rather than accumulate generic restrictions (Kane, 2013).

Assessment policies should consequently differentiate permitted support during learning from the conditions under which certification evidence is collected. A lecturer may quite properly encourage students to practise explanations with an AI tutor, while requiring an independent assessment of core reasoning. Conversely, a professional task may deliberately demand the responsible use of an AI system, in which case prohibiting all assistance would misrepresent the practice being assessed. Neither approach should be assumed superior without considering qualification level, construct relevance, accessibility, moderation and the claims eventually made to employers or professional bodies. Programmes need a documented balance between unaided and tool-mediated achievements, with independent demonstrations located where mistakes could affect public safety or the meaning of the award.

6.8 Academic integrity, authorship, AI disclosure, plagiarism, and limitations of detection systems

Academic integrity concerns honest participation in a scholarly community whose assessments, research and qualifications depend upon truthful representation. Synthetic audio, video and images can also impersonate participants or fabricate apparent practical evidence, so authorship checks cannot be confined to machine-written prose. AI alters the means by which assistance can be obtained but does not abolish the need to distinguish students' contributions from those of other people and systems. Traditional plagiarism, unauthorised collaboration, contract cheating, fabrication and misrepresentation remain relevant, while new cases arise when students present generated analysis as independent thought or conceal an AI-assisted procedure that assessment rules expressly prohibit. Yet the use of AI is not automatically plagiarism in the ordinary sense of unattributed human-source copying. Misconduct should be defined against the relevant rule, intellectual claim and evidence, rather than through imprecise labels (Miao & Holmes, 2023; Weber-Wulff et al., 2023).

The distinction matters for fair teaching and disciplinary decisions. AI-based grammar correction may be compatible with an assignment evaluating historical argument but inappropriate for an examination of unaided translation, depending on the announced criteria. Generating research questions may be permissible, while inventing interview data or sources is not. Use of a model to refactor working code may be legitimate in a software engineering task whose outcome includes tool evaluation; it could invalidate an introductory programming exercise assessing whether a learner can write and explain elementary logic. Students require prospective, task-specific instructions because general institutional statements cannot resolve such distinctions. HEC's 2025 guidance expressly addresses prohibited, permitted-with-attribution and encouraged uses, and recommends clear module assessment instructions rather than an unqualified blanket permission (HEC, 2025).

Authorship should be understood as accountable intellectual contribution, not the manual typing of every word. In academic research, the author is responsible for the truthfulness of representations, choice and interpretation of sources, analytical decisions and the final claims advanced. In student work, the exact expected contribution depends on the outcome being assessed. A student may legitimately use an aid while still performing the core reasoning, or may supply minimal direction to obtain a complete artefact that conveys little evidence of their competence. Institutions should require clear statements of the nature and extent of assistance, attribution where relevant, and a capacity to explain or defend the submitted work. An AI system should not be treated as a scholarly author capable of accepting responsibility, consent or institutional accountability (UNESCO, 2021; Miao & Holmes, 2023).

A proportionate disclosure record identifies the tool and, where relevant, version, the stages at which it was used, the purpose of assistance, any substantive content retained and the human verification conducted. Demanding complete transcripts for every incidental grammar suggestion would create excessive data collection and burden; accepting "AI used" without identifying material contributions may disclose too little. Institutions can set graduated expectations based on stakes, subject matter and privacy. HEC's Appendix IV provides an official AI-use declaration form covering categories of assistance and accountability. The original model in Table 6.2 offers a concise course-level supplement and should not replace the official form where that form applies (HEC, 2025).

Automated AI-text detection does not provide a defensible shortcut to authorship. Weber-Wulff et al. (2023) tested multiple detection systems and found substantial limitations in accuracy, reliability and resistance to paraphrasing. Liang et al. (2023) documented serious false-positive problems for non-native English writing samples in the tools they tested. Those studies do not prove that every later detector performs identically; they establish that high-stakes use requires current independent validation under relevant language and task conditions. A detector score cannot, by itself, establish that a particular student has committed misconduct. Equating a vendor's probability or category label with proof risks both false accusation and the overlooking of actual breaches.

A procedurally fair misconduct investigation should begin with a concrete potential rule breach, preserve relevant evidence securely, examine ordinary alternative explanations, invite the student to respond and obtain a reasoned decision from an authorised person. A lecturer may discuss the student's work or request an appropriately designed demonstration where institutional rules permit, but must not improvise a punitive oral examination solely because an essay "sounds like AI." Appeals, disability accommodations, language differences, data minimisation and records retention all matter. The academic question is what was authorised and what can be established, not whether suspicion can be made to feel persuasive. Detection may at most constitute a limited investigative signal under policy, not a self-authenticating verdict.

A disclosure process should be proportionate, understandable and prospective. For example, an assessment brief may state that generative AI can be used to develop a preliminary outline and language-edit a report, but not to invent data, sources or substantive findings. The student could then declare the service used, the nature of the assistance, the passages or decisions materially affected and the checks performed, while affirming responsibility for the submitted argument. Where no AI was used, a simple declaration may be appropriate if required under the institution's rules. A department should explain that disclosure of permitted assistance is not itself an allegation and should not trigger an automatic reduction in marks. The purpose is to make the academic contribution intelligible under the stated conditions, not to collect extensive personal activity records without educational justification.

Misconduct procedures require an especially disciplined distinction between suspicion, enquiry and finding. Unusual textual fluency, irregular draft histories, a disputed reference or a model-generated similarity flag may warrant an initial conversation, but none is intrinsically a determination of unauthorised assistance. An authorised reviewer should identify the actual rule in force when the work was completed, request only relevant evidence, provide meaningful opportunity to respond and explain the reasons for any conclusion. Accommodation and appeal procedures should be preserved. These safeguards protect academic integrity as well as students: a university that cannot distinguish a verified breach from an unexplained anomaly undermines confidence in its own decisions (Liang et al., 2023; Weber-Wulff et al., 2023).

HEC's 2025 guideline makes a notably cautious reference to a general 20–30% range for generated content, explicitly presenting it as discussion guidance rather than a strict rule and acknowledging imperfect detectors. Because "percentage of AI content" lacks a stable, reliable measurement basis across tasks and systems, institutions should not convert that passage into an automatic disciplinary threshold. A defensible policy instead specifies allowed functions and required individual contributions, which can be examined through assessment design and evidence. This interpretation respects the document's stated qualification while protecting procedural fairness and the validity of educational judgement (HEC, 2025; Liang et al., 2023; Weber-Wulff et al., 2023).

Table 6.2

Task-level AI-use disclosure model

Declaration field	Information to record	Why the field matters
Task and AI condition	Module, assessment, announced permissions	Tests use against the actual requirement
Tool identity	Service/model and version if reasonably available	Supports traceability without implying a fixed model
Nature of assistance	Ideation, translation, editing, coding, analysis, drafting or other	Distinguishes technical help from substantive production
Verification and human contribution	Sources checked, calculations repeated, decisions retained or changed	Shows what the declarant can responsibly defend
Protected information	Confirmation that applicable data rules were followed	Identifies confidentiality and permission risks
Accountability statement	Confirmation of accuracy, permitted use and own contribution	Preserves human responsibility and due process

Note. Author's proposed condensed teaching and assessment template, intended to complement HEC's official Appendix IV AI-use declaration, not replace it. Prompts or transcripts should be collected only when justified and lawfully handled.

6.9 Assessment redesign: Oral defence, portfolios, supervised performance, process evidence, and workplace-based assessment

Assessment redesign should begin with a map of vulnerability: not merely whether a generative model can produce a plausible answer, but whether the task can still establish the intended knowledge and judgement when permitted or unauthorised AI assistance is realistically available. The design team identifies what the graduate must do independently, what tool-mediated practice is legitimate in the profession and which evidence sources can support each inference. It then chooses an appropriate combination of assessment occasions, feedback, authentication and moderation. The goal is a coherent assessment system, not an arms race between student tools and increasingly burdensome restrictions (Kane, 2013; Miao & Holmes, 2023).

Oral defence can make reasoning visible when it tests explanation, transfer and response to probing counterexamples. It is particularly useful after a dissertation, policy report, design artefact or complex project. Yet oral questioning is not automatically fair or reliable. Examiners need an assessment blueprint, agreed criteria, realistic time allocations, disability accommodations, records, moderation and preparation that does not reward verbal fluency irrelevant to the construct. A short defence should test selected material in the submitted work, not become an unstructured search for inconsistencies. For large cohorts, sampling, structured group-based procedures or targeted high-stakes defences may be more sustainable than conducting lengthy interviews for every minor assignment.

Portfolios can document progression and make feedback use visible across time. Strong examples combine initial work, substantive revisions, reflection on choices, external feedback and a final defended artefact. Their weakness is that accumulated documents are not automatically authentic or sufficiently challenging. Institutions should specify milestone verification, supervisor involvement, criteria for independent contributions and retention of appropriate audit material. In digitally mediated programmes, portfolios can be accessible and longitudinal, but they require workable systems, clear ownership of evidence and arrangements for students with limited connectivity. A portfolio is valuable when it supports a developmental inference, not because the platform can store many files.

Supervised performance remains essential where practical competence has a safety, ethical or professional dimension. Clinical observation, laboratory operation, teacher practice, interpreting an unfamiliar dataset under controlled conditions and the live defence of software behaviour can each supply evidence inaccessible from a polished take-home report alone. Supervision must be proportionate to risk, appropriately staffed and consistent with professional regulation. In some disciplines it may be useful to permit AI during one stage and prohibit it during another: a student designs an engineering solution with a permitted assistant, then independently calculates a safety-critical quantity and defends the assumptions. This differentiation permits technological literacy without replacing foundational capability.

Workplace-based assessment can situate judgement within professional complexity, but an employer's positive comment does not replace a calibrated academic standard. Placement supervisors need clear descriptors, conflict-of-interest protections, documentation and appropriate university moderation. Where AI is embedded in the workplace, the programme should distinguish the student's actions from system-generated suggestions, the organisation's final decision and other team members' contributions. Small enterprises or rural placements may lack sophisticated monitoring systems, so evidence should focus on observable decisions, supervisor testimony, verified products and reflective explanation rather than prescribing costly software as the only acceptable record.

Table 6.3 translates this principle into an assessment-redesign instrument. No modality guarantees integrity, and no universally optimal mix is implied. The matrix should be completed alongside an outcome map, accessibility review and

workload appraisal. A well-designed revision may combine a brief AI-supported draft, an independently prepared analysis of unfamiliar counter-evidence and a moderated oral or written defence. The evidentiary strength emerges from convergence among performances, not from adding rituals that do not test the relevant competence.

Table 6.3

Assessment-redesign matrix for AI-rich learning environments

Learning claim and task	Permissible AI condition	Supplementary evidence	Review of validity and fairness
Explain foundational reasoning	Restricted or none during final demonstration	Supervised unfamiliar problem and explanation	Ensure accessibility and genuine conceptual demand
Develop research-informed argument	Bounded source discovery and editing with disclosure	Annotated sources, evidence checks and sampled defence	Do not mistake polished language for analysis
Make professional recommendation	Responsible AI assistance may be required	Decision record, alternatives, risk defence	Score judgement and verification, not output fluency
Demonstrate technical practice	Conditional tool use in designated stages	Observed performance, testing and error recovery	Protect safety and separate individual contribution
Produce creative or design work	Task-defined ideation or generation	Iterative portfolio, critique and process commentary	Credit original choices and contextual merit
Complete workplace placement	Rules consistent with host and programme	Supervisor observation, verified artefact and viva	Address privacy, power and consistency across sites

Note. Original proposed curriculum and assessment tool by the author. Its application requires programme-level alignment, moderation, access provisions and current regulatory verification.

A worked assessment example illustrates the necessary sequence of decisions. In a hypothetical upper-level public-policy module, the intended outcome is to appraise competing policy interventions using valid evidence and make a defensible recommendation. Students first produce a short independent interpretation of an unfamiliar official data extract under appropriate supervised conditions. They then prepare a longer evidence-informed policy memorandum with declared, bounded AI assistance for idea generation and editing. Finally, they defend a selected inference against new counter-evidence in a structured ten-minute oral or equivalent accessible written examination. The independent interpretation tests foundational analytical capacity; the memorandum tests research synthesis and professional communication; the defence tests transfer, uncertainty and the student's ownership of consequential judgement. No single component is described as a complete measure of competence, and each requires an appropriate marking rubric and moderation.

The assessment blueprint could allocate criteria to evidence quality, analytical reasoning, treatment of alternative explanations and accountable recommendation rather than rewarding stylistic fluency. It would also indicate which evidence is individually attributable, which stages permit assistance, what accessibility routes are available, how second marking or calibration occurs, and how suspected breaches are considered separately from academic grading. An institution with limited staff capacity might replace a full oral examination of every formative task with moderated short defences focused on high-stakes outcomes, supplemented by properly recorded supervised work. The objective is not to impose a costly universal procedure but to ensure that the total assessment system supports the specific qualification claims. In professional programmes, any required practical observations or external licensing obligations remain additional constraints (Biggs, 1996; Kane, 2013).

6.10 Algorithmic bias, data protection, transparency, explainability, and learner rights

AI may perpetuate unequal treatment through biased training data, unsuitable outcome definitions, inaccessible interfaces, inappropriate proxy variables or unrepresentative testing. Bias is not exclusively a property of an algorithm: institutional practices can be discriminatory before automation, and automated systems can reproduce or amplify existing patterns. A learner-support model that relies heavily on frequent platform login might misidentify employed students who study offline, students sharing devices or those with accessibility constraints. An automated scoring system trained on dominant forms of academic English might disadvantage legitimate linguistic variation. Governance must examine purpose, data, model behaviour and the human decisions that follow, not simply whether a vendor describes its system as fair (Liang et al., 2023; UNESCO, 2021).

Fairness cannot be established through a single aggregate accuracy statistic. Performance should be tested for the populations and subgroups materially affected, using lawful and ethically appropriate data, and interpreted with awareness of sample limitations. In small cohorts, subgroup analysis can become unreliable or threaten confidentiality, so qualitative error review, stakeholder participation and conservative decision rules may be more useful than invented numerical precision. Educational equity also involves distribution of benefit: a system may function accurately for those

who can afford a premium subscription while excluding other learners. Institutions should establish equivalent academic routes where AI access is not essential to the assessed outcome and explain when access is genuinely a curricular requirement.

Rwanda's Law No. 058/2021 of 13 October 2021 relating to the protection of personal data and privacy imposes legal responsibilities relevant to processing student and staff information. The competent supervisory structure and particular duties must be interpreted from the current law and applicable implementing arrangements, not guessed from a vendor's contract. Educational use may involve names, academic results, disability accommodations, attendance, research records or disciplinary concerns. Before submitting such information to a third-party model, the university must identify the lawful basis and necessity of processing, data categories, retention, access, transfers, security and applicable rights. A lecturer's convenience does not confer independent authority to disclose confidential student work to an unapproved service (Republic of Rwanda, 2021; Data Protection and Privacy Office, n.d.).

Purpose limitation and data minimisation should be practical design principles. A formative tutor may need to know which concept a learner is studying without knowing the learner's home address, health history or disciplinary file. A system generating a generic case can often use synthetic data rather than a real client's records. A reviewer analysing academic standards should not upload unredacted candidate lists or confidential accreditation submissions to an external system absent authorisation and adequate safeguards. Procurement agreements must address model training on institutional inputs, retention and deletion, subprocessors, cross-border data movement, incident response, audit access and meaningful contract termination. These matters are integral to educational quality because a breach can undermine learner trust and continuity.

Explainability has several levels that should not be conflated. A lecturer may explain how a tool is used in an assessment; the institution may explain why a particular system was approved; a model developer may offer information about design and limitations; and an affected student may reasonably ask why an adverse decision was made. The university does not need to provide a technically impossible complete reconstruction of every model parameter to give a clear account of its own decision. It must, however, disclose material automated involvement where required, provide intelligible reasons for consequential academic actions and ensure an authorised human can reconsider disputed evidence. An explanation that merely repeats an opaque AI output is not adequate accountability (NIST, 2023; UNESCO, 2021).

Student rights are especially important where systems rank applicants, infer behaviour, detect suspected misconduct or determine progression. A low-stakes tutoring recommendation may be reversible; an unjustified accusation, adverse grade or admission decision can have durable effects. Risk should reflect severity, scale, vulnerability, likelihood, reversibility and the learner's capacity to contest the action. Proportionate safeguards include advance notice, meaningful choice when feasible, human review, documented reasons, correction processes and accessible complaint pathways. Technological innovation should never convert administrative convenience into an exemption from academic fairness.

Language and disability accessibility must be assessed at the point of actual use. A multilingual chatbot can produce a grammatically plausible answer while misunderstanding a technical concept in Kinyarwanda or omitting context from a locally relevant example. A speech-enabled tutor may improve access for some learners but fail students using screen readers, noisy shared environments or inexpensive devices. Institutions should therefore involve representative users, check real task completion, provide correction routes and retain alternative forms of participation where equivalent learning can be achieved without the tool. This is also a question of epistemic fairness: evaluation should not mistake facility in a model's dominant language or conversational style for superior disciplinary judgement. Appropriate tests should examine whether differences in performance arise from educationally relevant competence or from tool and access conditions (UNESCO, 2021; Liang et al., 2023).

Institutional information governance should account for the data life cycle rather than focus only on the moment of upload. Relevant questions include whether personal data are collected lawfully, whether identifiable inputs are used for model improvement, who can retrieve conversations, whether subprocessors or cross-border transfers are involved, how long logs persist, how deletion requests are executed and how a serious incident is escalated. A contract promising confidentiality is weak protection if staff continue to disclose identifiable coursework through unapproved personal accounts. Institutions need both technical measures and understandable staff guidance, coupled with auditing proportionate to the sensitivity and scale of use. A student's consent, where relied on, should not be treated as freely given if the only practical route to an essential course activity requires surrendering unnecessary data (Republic of Rwanda, 2021).

6.11 Academic labour, faculty professional development, institutional procurement, and technological dependence

AI is frequently presented as a productivity instrument, but the relevant institutional question is where work is displaced, intensified or newly created. Rapid drafting can reduce some preparation time while increasing obligations for verification, assessment redesign, student consultation, disciplinary investigations and vendor oversight. Automated

first-pass feedback may release time for more demanding discussion, or become a pretext for expanding cohorts without preserving lecturer availability. Economic claims about efficiency should therefore be tested against full work processes, not merely against the time a machine requires to produce a text. A university should account for academic, technical, legal, accessibility and information-security labour across the whole system.

Academic labour also carries professional and epistemic responsibilities that cannot be delegated by contract. Lecturers must choose what is educationally worth teaching, protect disciplinary standards, interpret students' reasoning and exercise judgement where values and evidence conflict. An institution that uses AI to generate most teaching material while removing specialist verification may reduce visible labour but increase intellectual and safety risk. Conversely, blanket refusal to use suitable tools may squander opportunities for accessibility, multilingual explanations or rapid formative support. The appropriate criterion is whether the technology improves the institution's capacity to fulfil its educational commitments without displacing human accountability (Miao & Cukurova, 2024; Zawacki-Richter et al., 2019).

Procurement is a curriculum and quality-assurance decision whenever a system affects teaching, learning, assessment or academic records. A responsible approval file should contain an educational case, consideration of alternative approaches, lawful data-processing assessment, security review, accessibility test, rights and intellectual-property terms, cost analysis, evaluation of vendor claims, pilot plan, assigned system owner and an exit strategy. A free or discounted trial can create significant future costs if data, student workflows or course material become difficult to migrate. An institution should identify whether it can export records in usable forms, retain a viable non-AI route and continue teaching if service access is interrupted or prices rise.

Technological dependence has a geopolitical and economic dimension. Cloud services can be concentrated among a small number of providers, with foreign currency charges, changing national restrictions, shifting terms and outages affecting local users. Hosting a smaller model locally may improve control over some data and continuity, but it creates its own costs in hardware, energy, maintenance, security, specialised staff and model evaluation. “Local” is not synonymous with secure or unbiased; “open source” does not, without examining the actual licence and available components, establish unrestricted use or reproducibility. Decision-makers should compare feasible operating arrangements through total cost, governance rights, safety and educational performance rather than slogans about sovereignty.

Environmental consequences are also relevant to procurement and curriculum decisions. AI services use energy and computing resources, and the scale of impacts depends on infrastructure, deployment and use patterns. Precise attribution of environmental effects to a particular student prompt is often methodologically uncertain. Universities can nevertheless avoid gratuitous high-compute use, demand reasonable information from suppliers, optimise unnecessary repetition and compare alternatives without presenting unverified emissions calculations as evidence. The broader task is responsible stewardship of educational, financial and environmental resources consistent with the institution's mission and documented constraints (UNESCO, 2021; Tabassi, 2023).

Professional development cannot be an isolated induction webinar. Institutions need continuing communities of practice, disciplinary mentoring, confidential support for problematic use cases and recognition that staff require time to redesign curricula. Academic managers should establish workload allowances for assessment review and quality verification rather than expecting lecturers to absorb new governance burdens without resources. Staff should have clear channels for refusing unsafe use, reporting system failures and requesting reconsideration of tools imposed centrally. Sustainable AI capability is a coordinated organisational function, not an accumulation of individual subscriptions.

A distinct technical risk arises when an AI system can retrieve external content or execute tools. Instructions embedded in a webpage, assignment, repository document or message may attempt to redirect the system's behaviour, including toward disclosure of information or unauthorised changes. This form of prompt injection is not resolved merely by asking the model to ignore malicious text. Universities should implement least-privilege access, separate untrusted content from authoritative instructions, require human approval for consequential actions, monitor transactions and test recovery procedures. An AI agent that can post grades or contact students should be evaluated with the same seriousness as other software holding comparable privileges, with additional attention to unpredictable output and instruction-following failure. The minimum defensible position is that no system receives more data or action authority than its approved function requires (Autio et al., 2024; Tabassi, 2023).

Financial appraisal should similarly differentiate a tool's purchase price from the total cost of reliable educational provision. This includes staff verification time, API or subscription usage, cloud storage, accessibility support, incident response, data processing, updating assessment policies, curriculum redevelopment and exit or migration costs. Institutions should examine budget scenarios for larger cohorts, currency volatility and loss of subsidised pricing, and assess whether an outage would interrupt essential instruction or regulated examinations. Low-cost experimentation may be appropriate for reversible formative functions; procurement affecting programme-level assessment or long-lived academic records requires a more durable evidence base and decision trail. Cost-effectiveness cannot be inferred merely from the number of texts generated per minute.

6.12 Governance of AI-supported programme evaluation and accreditation

External review is especially vulnerable to confident automation because accreditation dossiers contain structured claims, repeated sections, evidence registers and administrative language that appear well suited to machine summarisation. AI may help identify missing fields, compare programme specifications, flag inconsistent module codes or prepare a preliminary index of evidence. Yet the reviewer must establish that the tool correctly read the current document, distinguished versions, understood regulatory applicability and did not invent deficiencies or compliance findings. A system can produce a persuasive summary while concealing missing annexes, inappropriate reliance on obsolete rules, or incorrect inferences about readiness (Sangwa, 2026; HEC, 2025).

The governing distinction is between assistance with information management and authority to make a judgement. In Rwanda, the applicable legal instruments determine HEC's powers and the authorised decision route. A professional reviewer's role is to examine evidence and make findings or recommendations within their mandate; a tool acquires no regulatory competence through its usefulness. AI-generated scores should not replace independent consideration of curriculum substance, qualification level, available staff, assessment validity, student protection or programme sustainability. A legally important deficiency must be tied to an operative norm and genuine evidence, while missing evidence should be described as an uncertainty rather than automatically as proof of substantive failure (Republic of Rwanda, 2026; Sangwa, 2026).

The author's Evidence-Traceable Accreditation Judgement (ETAJ) model is useful as a bounded reasoning aid. For every material conclusion, an evaluator specifies the normative source, the institutional claim, the original evidence, steps taken to verify it, the evaluative warrant, counter-evidence, remaining uncertainty and proportionate finding. In an AI-assisted review, the record adds the system's identity, the exact task delegated, the documents accessed, the output checked and the human who accepted or rejected it. This extension is consistent with the overarching Integrated Evidence-Traceable Quality Governance (IETQG) framework because institutional capability, programme delivery and regulatory judgement remain connected through inspectable evidence. Neither model is an official HEC procedure, a validated predictive tool or permission to substitute private judgement for competent authority.

For an illustrative curriculum review, imagine that a model reports that the module “Research Design” is absent. The reviewer consults the actual programme register and finds that equivalent research competencies are deliberately distributed across two modules and assessed in a final project. A title-matching algorithm has identified an apparent gap, but the relevant academic question is whether the competencies are truly covered at the intended level. Conversely, an automated comparison may discover a real duplicate code or contradictory credit total that human reviewers initially overlooked. The model is then useful as a prompt for verification, not as the source of the final finding. The authoritative evidence remains the checked curriculum and applicable standard.

Confidentiality and independence require particular safeguards. Institutional submissions may contain personal records, unpublished research, assessment instruments, commercial agreements, private staff contracts and sensitive financial information. Uploading the dossier to an unapproved public system may breach confidentiality even if the evaluator's intention is to improve efficiency. Review bodies should define permitted systems, lawful access, retention, publication restrictions, conflicts of interest and records of AI assistance. Where AI is used in report drafting, the reviewer should confirm that every stated observation is supported, every cited rule applies and every recommendation reflects their own professional evaluation. The reviewer must be able to defend the result without invoking the software's authority.

HEC's 2025 Appendix III supplies a concrete official assessment tool for AI integration, with seven components and a stated 84-point total. It describes 'ready with minor corrections' at a score of at least 67, subject to attaining at least half in every component, and 'conditional' at a score of at least 42, also subject to the component minima; failure of the stated minima can lead to 'not yet ready'. The document also describes 67/84 as 'at least 80%', although the mathematical proportion is approximately 79.76%; borderline application therefore calls for authorised interpretation of the published rule rather than an evaluator's silent numerical correction. Institutions and reviewers should consult the exact current wording and component-level decision conditions rather than substitute the author's new scoring framework. Scores cannot be interpreted independently of stated thresholds, required evidence or safety-critical deficiencies, and evidence of adopting a system does not demonstrate improved graduate outcomes. Table 6.4 therefore supplements, but does not reproduce or supersede, HEC's assessment instrument through a claim-evidence-verification record for AI-supported review (HEC, 2025).

Table 6.4

Evidence-traceable AI-assisted evaluation record

Element	Reviewer entry	Verification standard
Applicable requirement	Authority, provision, version and scope	Norm is operative and applicable to the claim
Institutional claim	Precisely what the institution represents	Do not infer performance from policy titles

Element	Reviewer entry	Verification standard
AI assistance	Approved tool, task and documents accessed	No unauthorised confidential disclosure
Source evidence	Original record, date and responsible owner	Authenticity, currency, completeness and relevance
Independent test	Human inspection, demonstration or corroboration	AI output never stands as the sole evidence
Finding and uncertainty	Warrant, counter-evidence, limitation and action	Reasonable, reviewable and proportionate conclusion

Note. Original application of ETAJ principles to AI-assisted review by the author. This worksheet is a non-regulatory supplement and has not been empirically validated.

The quality of an AI-assisted accreditation finding depends upon reviewer calibration as well as confidentiality. Two competent evaluators examining the same programme should be able to identify the same governing source and substantial evidence, even if they reasonably differ about the weight of a borderline deficiency. A review team can improve consistency by testing sample cases before a visit, recording how ambiguities were resolved and reserving unresolved statutory questions for the legally competent authority. A tool may assist document indexing or arithmetic verification, but a finding should never rest on a generated statement that cannot be traced to a page, observed demonstration or other original source. Review records should also distinguish failure to produce requested evidence from positive evidence that a programme does not possess the requisite capability. This prevents uncertainty being converted into an unjustified adverse fact.

For example, suppose a hypothetical institutional dossier contains an AI policy, a screenshot of a chatbot integrated into a learning platform and a list of staff who attended orientation. Those records might support a finding that policy and initial deployment exist, but they cannot establish safe data handling, assessment validity, lecturer competence or an effective complaints process. The evaluator would seek a lawfully accessible system demonstration, a sample of actual module rules, appropriately redacted assessment evidence, staff responses to realistic failure scenarios and proof that incidents are reviewed. The resulting judgement should describe precisely which claims are supported, which remain unverified and which require corrective action. This differentiated approach is more professionally defensible than either praising technological visibility or rejecting the programme on an unverified presumption of risk (HEC, 2025; Sangwa, 2026).

6.13 Institutional AI policy development, accountability, risk classification, and policy revision

Effective policy must be specific enough to guide real decisions yet adaptable enough to survive technology changes. It should state educational purposes, lawful authority, scope, roles, acceptable and prohibited practices, task-level disclosure, permitted data handling, procurement approval, research and assessment controls, student complaints, staff development, incident reporting and periodic review. A policy that merely celebrates innovation lacks practical limits; one that prohibits every form of machine assistance may be neither enforceable nor educationally defensible. The more useful approach is to classify applications by purpose and consequence, define minimum safeguards and require explicit approval for significant departures (Miao & Holmes, 2023; NIST, 2023).

Rwanda's 2025 HEC guidance is unusually operational on these matters. It addresses a dedicated AI function, cross-functional committee, annual monitoring, reviewed literacy module, staff development, assessment-specific AI-use instructions, governance policies, tool appraisal and declarations. The proper institutional response is to map current provisions to existing committee powers, appoint accountable staff and produce evidence that the required functions operate. A small university may implement coordination through an existing teaching-and-learning or ICT unit where consistent with HEC's instrument; naming a new centre without staff, decision rights or records would demonstrate little capability. The HEC Appendix III instrument should be used on its own terms where applicable, while this chapter's proposed risk form supports local implementation and critical review (HEC, 2025).

A useful risk classification considers at least six dimensions: educational consequence, data sensitivity, degree of autonomous action, affected population and equity, reversibility, and external dependencies. Routine generation of fictional practice questions from public information may require ordinary teacher verification. A student-facing tutor processing identifiable responses requires privacy and pedagogical controls. Automated summative grading, admissions ranking, disciplinary screening, clinical advice or autonomous access to student records require much stronger validation and governance, and some uses may be inappropriate regardless of proposed benefits. Risk should be reviewed at the level of the actual workflow rather than assigned forever to a brand or product category.

The NIST AI Risk Management Framework distinguishes governance, mapping, measurement and management as interdependent functions. This provides a useful voluntary benchmark for institutional procedure: establish accountable decision-making; understand the context and potential harm; test the system and relevant claims; and manage or discontinue unacceptable risks. Its generative-AI profile broadens attention to risks that arise from generative systems, but neither NIST instrument has the force of Rwanda's domestic law simply because it is technically sophisticated.

International reference points should assist local judgement, not obscure the authority hierarchy (Tabassi, 2023; Autio et al., 2024).

Operational policy needs a living inventory of approved and unapproved use cases, versioned assessment rules and records of incidents or near misses. Triggers for extraordinary review might include a model replacement, substantial change to vendor terms, a security event, discriminatory outcome, new HEC instruction, altered professional standard or evidence of repeated unreliable assessment. The institution should document whether the action is to retrain staff, correct content, amend an assessment, suspend an integration or terminate a contract. Annual review, while important, is not sufficient when a serious incident requires immediate containment. Equally, frequent policy edits that are not communicated can create unfair retrospective expectations for students.

Table 6.5 translates differentiated risk review into a qualitative, non-compensatory appraisal. A defensible policy should also identify the authority to grant exceptions and the conditions for doing so. Reasonable accessibility accommodations, secure professional simulations or approved research protocols may justify departures from ordinary tool conditions; secret exceptions awarded on personal relationships would undermine fairness. Decisions should distinguish mandatory legal or HEC requirements from institutional discretion and disciplinary judgement. A published policy is evidence of intention. Capability is stronger when the university demonstrates working approvals, assessments with clear conditions, staff competence, documented objections and corrections, and improved practice after review.

Table 6.5

Risk-based AI-tool evaluation form for universities

Review domain	Evidence and test required	Decision or escalation trigger
Educational purpose	Outcome, alternatives, pedagogical rationale and success evidence	Reject if no legitimate educational function
Legal/data protection	Processing basis, retention, transfers, access, security and contracts	Do not deploy with unresolved material legal risk
Validity and reliability	Representative tests, error analysis, known limits and human fallback	Escalate consequential unreliable outputs
Equity and accessibility	Language/disability testing, cost, devices, realistic alternatives	Correct exclusion or deny deployment
Autonomy and control	Permissions, audit logs, intervention rights and rollback	Prohibit uncontrolled high-impact actions
Institutional capability	Named owner, trained staff, support, full costs and exit route	Defer if continuing operation is not credible
Review plan	Metrics, complaints, incident thresholds and reevaluation date	Suspend or reassess after significant change

Note. Original risk-based professional form proposed by the author. Entries are qualitative and deliberately non-additive: an unresolved critical legal or safety risk cannot be offset by a favourable total elsewhere.

An operational institutional AI policy should define decision rights, not merely committee membership. The governing body or delegated academic authority approves principles and tolerable institutional risk within its mandate; senate or equivalent academic leadership protects curriculum standards and academic regulations; the institutional data-protection and security functions assess lawful processing and technical risk; programme teams determine educational fitness and assessment conditions; internal quality assurance verifies evidence and corrective action; and an authorised appeals function reconsiders adverse individual decisions. These are functional responsibilities, not a universal legal organogram. Their precise allocation must be reconciled with an institution's statutes, Rwanda's current HEC instruments and relevant sector-specific requirements. Where one employee performs several roles, the institution should document independent challenge and escalation for decisions that materially affect students. Table 6.6 sets out corresponding institutional functions and the evidence by which an internal auditor could test their operation.

Policy implementation also requires an explicit sequence for exception handling. A lecturer seeking an unapproved tool for sensitive coursework should identify the instructional need and the data involved before any upload. Academic and technical reviewers should determine whether the objective can be achieved with a lower-risk service, synthetic data or an alternative learning activity. If a proposal remains appropriate, the designated authority records conditions, limits, monitoring responsibilities and the review date. Where an incident subsequently occurs, the university should preserve relevant records, contain the harm, inform affected persons where required, provide an academic remedy and determine whether the tool should remain in service. The reasoning and response should be proportionate to consequence, with student rights and programme continuity considered together. These processes turn a policy from a statement of aspiration into reviewable institutional capability (HEC, 2025; Republic of Rwanda, 2021).

Table 6.6

Decision rights and evidence of functioning institutional AI governance

Responsible function	Decision or verification responsibility	Record demonstrating enactment
University governing authority	Authorize institutional risk policy and resourcing within its legal competence	Approved policy, resourcing decision, risk escalation
Senate or academic board	Safeguard outcomes, award rules and academic assessment standards	Versioned regulations, assessment blueprint and moderation findings
Programme and teaching teams	Specify permissible AI use and verify educational fitness	Module instructions, learning samples and corrective revisions
AI/ICT and data-protection functions	Appraise vendors, data permissions, technical control and incident response	Tool register, impact assessment, test and incident log
Quality assurance unit	Verify implementation and follow through on identified deficiencies	Evidence register, review findings and action closure
Authorised complaints or appeals function	Provide impartial procedural review of consequential decisions	Recorded reasons, decisions and academic remedies

Note. Original proposed decision-rights matrix by the author. The responsible organ and legal powers must be adapted to institutional statutes and applicable HEC instruments; roles may be combined only with adequate safeguards against conflicts of interest.

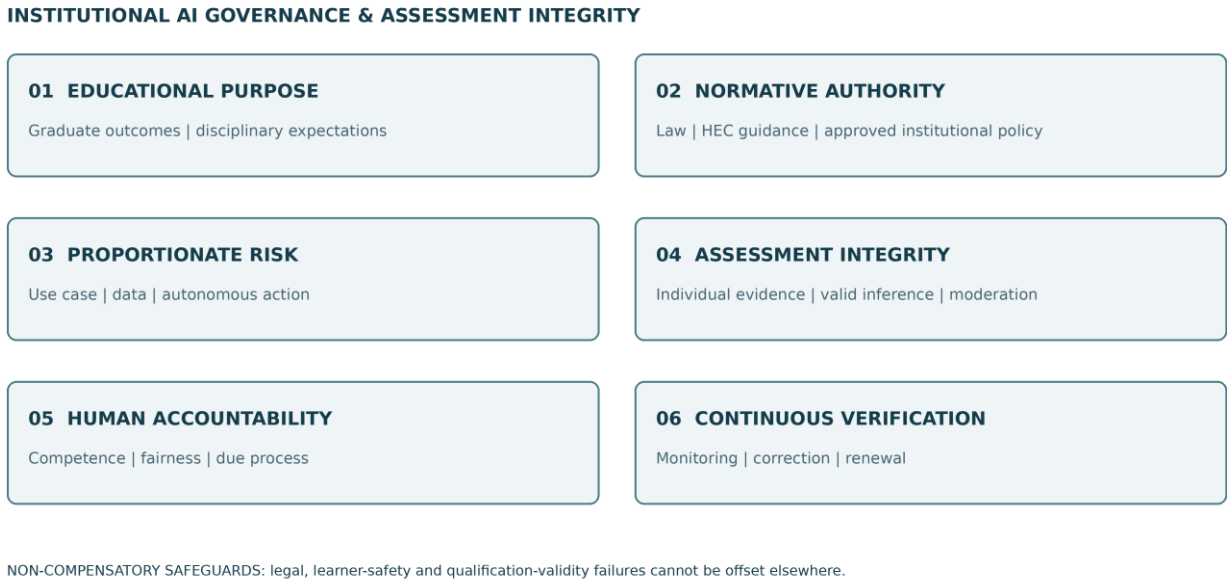
6.14 Towards an evidence-based model of responsible AI-enabled higher education

The preceding analysis identifies a recurrent failure in institutional AI debates: attention often begins with the technology and ends with a statement of ethical principles, leaving the educational judgement insufficiently specified. An integrated model should proceed in the opposite direction. Begin with the educational or institutional claim, identify its legal and ethical boundaries, select an appropriate function for AI, protect the evidence needed to evaluate achievement, secure human accountability and continue review after deployment. This is the proposed *Institutional AI Governance and Academic Assessment Integrity Framework* (IAG-AAIF). It is an interpretive and operational contribution, not an empirically tested prediction model or an official HEC instrument.

The framework contains six interacting controls. **Educational purpose and standards** establish what must be learnt and demonstrated. **Normative authority and institutional policy** define what is lawful, officially required and locally authorised. **Use-case suitability and proportional risk** determine whether the proposed AI assistance is educationally justified and governable. **Assessment and evidence integrity** protect the inference from student or institutional performance to an academic claim. **Human rights, competence and accountability** require meaningful oversight, accessibility, privacy and due process. **Continuous verification and adaptation** connect observed operation, incidents, feedback and updated sources to improvement. These controls are not six interchangeable points: serious deficiency in legality, learner safety or qualification validity cannot be cancelled by strong performance elsewhere.

Figure 6.2

Six mutually dependent controls in responsible AI-enabled quality governance



Note. Original IAG-AAIF conceptual model by the author. The connections express proposed institutional reasoning and iterative review, not empirically established causal pathways.

An implementation sequence can translate the controls into institutional work. Senior management and academic governance first reconcile existing law, HEC instruments, strategy and assessment regulations, documenting any ambiguity requiring authoritative clarification. Programme teams then review graduate outcomes and classify assessments by the individual competence they must demonstrate. The designated AI governance function compiles a use-case register, evaluates tools, coordinates privacy and security checks and confirms access alternatives. Teaching and learning staff test limited deployments, document student and lecturer experience, and amend assessment or support arrangements before scale-up. Internal QA subsequently verifies practice against the approved design, and senior academic authorities review exceptions, risks and evidence of improvement. Each handover should retain the relevant source, decision-maker, date and accountable owner.

Consider a hypothetical Rwandan bachelor's programme in business analytics. Its staff propose using an external AI service for formative explanations, AI-assisted production of client reports and automated scoring of final strategic recommendations. The first use might be permitted with source verification and an accessible alternative. The second could be taught as a professional competency if students document datasets, prompts, corrections and justified decisions, and can independently interpret unfamiliar evidence. The third raises a more serious validity and fairness problem. Without a sufficiently supported scoring interpretation, representative testing, contestability and meaningful faculty moderation, an automated grade could be educationally indefensible even if the software is technically available. The proper response is differentiated approval, not uniform enthusiasm or prohibition.

The institution's evidence register for this illustration would include the approved learning outcomes, assessment rules, authorised tool inventory, privacy decision, learning materials, lecturer training, student declaration process, moderated assessed samples, documented failures, action records and review dates. Such an evidence trail does not prove that graduates are more employable or that AI causally improved learning. It does allow reviewers to examine whether the university's claim that it governs AI responsibly has a reasonable basis. If students cannot defend their recommendations or staff cannot explain how automated scores were checked, the responsible authority must revisit the assessment and potentially suspend that use. The willingness to amend a failed implementation is a dimension of institutional maturity.

The proposed framework complements IETQG by locating AI inside institutional capability, programme assurance, evidence generation and continuing improvement. It complements ETAJ by making every material finding traceable to the governing criterion, examined record, verification procedure and evaluative warrant. The models do not remove professional disagreement; rather, they clarify where disagreement lies and what additional evidence might resolve it. Future scholarship should test feasibility, inter-rater consistency, consequences for workload, learner experience, educational validity and bias in different disciplines and African institutional contexts. Until those tests occur, stronger language about effectiveness would be unjustified.

The framework becomes usable only when deployed as an iterative governance process supported by evidence, not as a decorative diagram. A realistic initial cycle begins with an inventory of educational AI uses and an authority check, proceeds through programme-level assessment mapping and staff competence review, and then selects limited use cases for properly evaluated pilots. Decision-makers identify success and harm indicators in advance: verified learning against intended outcomes, independent transfer performance where relevant, accessibility, student complaints, integrity cases substantiated through fair procedure, staff workload and total cost. The pilot record should also document rejected

options and unexpected negative evidence. Subsequent expansion should be conditional on those findings, and adverse events should trigger review without waiting for the next annual reporting window. The point of the proposed sequence is to make changes to educational work intelligible and accountable, not to suggest a universal timetable for institutions of different size or maturity. Table 6.7 translates these requirements into a sequenced, auditable implementation process.

Table 6.7

Illustrative implementation and assurance roadmap for institutional AI adoption

Phase and primary concern	Expected output and evidence	Decision gate
1. Map authority and existing use	Regulatory provenance record, tool inventory, current assessment rules	Unresolved legal or critical security risk blocks use
2. Protect learning claims	Graduate outcome and assessment-condition map, identified independent demonstrations	Assessment inference remains educationally defensible
3. Appraise and select tools	Use-case rationale, privacy and accessibility tests, alternatives and owners	Approval is bounded to use, users and data
4. Pilot and verify	Representative learner tasks, independent performance, errors, appeals and workload	Expand only after material risks are controlled
5. Implement and monitor	Training artefacts, contracts, quality review, incident and correction records	Serious failures trigger immediate containment
6. Review and renew	Documented reassessment after policy, vendor, curriculum or evidence change	Continue, amend, suspend or discontinue with reasons

Note. Proposed institutional adaptation of the IAG-AAIF; it is not a prescribed HEC schedule or empirically validated intervention. Each institution should assign named owners, deadlines and evidence identifiers appropriate to its context.

The research agenda is correspondingly specific. Future studies could compare AI-assisted and non-assisted pathways using baseline-sensitive, delayed independent assessments; test whether review panels using the evidence-traceability worksheet reach more consistent and well-supported conclusions; and investigate whether policies disproportionately burden particular language groups, disability groups or resource-constrained students. Such work requires ethical governance, defensible samples, instruments capable of assessing intended constructs and transparent reporting of null or unfavourable findings. The present chapter contributes a conceptually integrated professional model and practical instruments; it does not claim that implementing them has already improved attainment, accreditation consistency or institutional resilience. This evidentiary limit is necessary if the proposed IAG-AAIF is to function as a serious researchable framework rather than an assertion of success (Deng et al., 2025; Chen & Cheung, 2025; Sangwa, 2026).

6.15 Conclusion: Protecting educational judgement in an AI-rich university

The central issue in AI-enabled higher education is not the presence of machines in academic work but the credibility of the claims universities make when machines participate in it. A curriculum must still represent worthwhile knowledge; teaching must still provide meaningful opportunities to learn; and assessment must still distinguish credible individual achievement from the mere production of plausible artefacts. Human oversight is not secured by placing a lecturer's name beside a machine-generated decision. It requires competence, authority, verification, sufficient time and a process by which errors and unfair outcomes can be corrected. Without those conditions, the language of responsible innovation becomes an administrative assurance unsupported by educational practice.

This chapter has distinguished varieties of AI, defended discipline-sensitive graduate competence, explained the inferential consequences of assistance for assessment validity, and proposed safeguards for authorship, privacy, bias, institutional procurement and accreditation review. It has also situated these safeguards in Rwanda's evolving policy and HEC environment without treating an international benchmark as national law or turning a suggested professional instrument into an official requirement. The IAG-AAIF, disclosure model, assessment-redesign matrix and risk-based review form translate the analysis into tools for discussion, testing and local adaptation. Their conceptual usefulness must not be mistaken for demonstrated improvement in student outcomes or independent empirical validation.

The practical implication is demanding but clear. Institutions should invest as much intellectual effort in deciding what evidence would justify an AI-supported academic judgement as they invest in acquiring the system that produces it. Universities serving students with diverse languages, resources and accessibility needs should be particularly cautious about rules that make affluent tool access a hidden criterion of achievement or unreliable text detection a proxy for integrity. Regulators and peer reviewers should request evidence of actual practice, lawful processing, relevant staff competence and fair remedies, while remaining alert to the uncertainty of predictive claims and the limits of technology

demonstrations. Responsible AI governance ultimately strengthens academic freedom and institutional credibility by making decisions open to reasoned scrutiny.

The argument also prepares the next question in the volume. Protecting academic standards cannot be achieved only through assessment design and technical controls; it depends upon whether all learners can meaningfully participate in teaching, obtain timely support and demonstrate attainment under equitable conditions. The following discussion of inclusive curriculum and student success therefore examines the learners for whom these AI-enabled arrangements must work, including those whom a formally sophisticated system might otherwise overlook.

References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364. https://doi.org/10.1007/BF00138871

Contractor, Z., & Reyes, G. (2026). Experimental evidence on the learning impact of generative AI (IZA Discussion Paper No. 18792). IZA@LISER Network. https://www.iza.org/index.php/publications/dp/18792/experimental-evidence-on-the-learning-impact-of-generative-ai

Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. Educational Research Review, 49, Article 100737. https://doi.org/10.1016/j.edurev.2025.100737

Data Protection and Privacy Office. (n.d.). *Who we are*. https://dpo.gov.rw/about-us/who-we-are

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Higher Education Council. (2025). *Guidelines and assessment tools for the virtual learning and the use of artificial intelligence in Rwanda's higher learning institutions*. https://www.hec.gov.rw/fileadmin/user_upload/HEC/Publications/Guidelines/GUIDELINES_ASSESSMENT_TOOLS_FOR_VL_and_AI_in_HLIs.pdf

Higher Education Council. (2026). *Guidelines and assessment tools for distance learning in Rwanda's higher learning institutions*. https://www.hec.gov.rw/fileadmin/user_upload/HEC/Publications/Guidelines/GUIDELINES_AND_ASSESSMENT_TOOLS_FOR_DISTANCE_LEARNING.pdf

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391105

Ministry of ICT and Innovation. (2023). *The national AI policy*. Government of Rwanda. https://www.minict.gov.rw/fileadmin/user_upload/minict_user_upload/Documents/Policies/Artificial_Intelligence_Policy.pdf

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*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Republic of Rwanda. (2021). *Law No. 058/2021 of 13/10/2021 relating to the protection of personal data and privacy*. *Official Gazette*, special issue of 15 October 2021. https://rwandalii.org/akn/rw/act/law/2021/58/eng@2021-10-15

Republic of Rwanda. (2026). *Presidential Order No. 044/01 of 24/07/2026 governing the Higher Education Council*. *Official Gazette*, special issue of 30 July 2026. https://www.hec.gov.rw/fileadmin/user_upload/HEC/Publications/Laws/Presidential_Order_n___04401_of_24072026_governing_the_Higher_Education_Council.pdf

Sangwa, S. (2026). *Governing quality in digitally transforming African higher education: An integrated framework for institutional capability, programme quality assurance, and accreditation judgement* [Doctoral thesis, Open Christian University]. https://doi.org/10.65655/ocur.m28

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

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