HEA | An tÚdarás um Ard-Oideachas
The Higher Education Authority

Generative AI in Higher Education Teaching & Learning

Principles for Ethical AI Adoption

Contributors

James O'Sullivan
Colin Lowry
Ross Woods
Tim Conlon

Acknowledgements

Alan Smeaton (Government of Ireland AI Advisory Council), Barry O'Sullivan (Government of Ireland AI Advisory Council), Susan Leavy (Government of Ireland AI Advisory Council), Caoimhe Hope (Department of Further and Higher Education, Research, Innovation and Science), Anne Ribault-O'Reilly (Department of Further and Higher Education, Research, Innovation and Science), Ana Rocha (Higher Education Authority), Imma Zoppi (Higher Education Authority), Paul O'Donovan (University College Cork), Joseph Feller (University College Cork), Danielle Duignan (AHEAD), Jim O'Mahony (Munster Technological University), Denise Mac Giolla Rí (Technological University of the Shannon), Justin Tonra (University of Galway), Brian Marrinan (Journey Partners), Clyde Hutchinson (Journey Partners), Mary-Claire Kennedy (University of Limerick), Maria Murphy (Maynooth University), Derek Dodd (Technological University Dublin), Pauline Rooney (Trinity College Dublin), Loretta Goff (University College Cork), Roisín Morris-Drennan (Quality and Qualifications Ireland), Liam Fogarty (University College Dublin), Barbara Whelan (Microsoft), Johanna Archbold (Atlantic Technological University), Áine Clarke (Ibec), Marie Clarke (University College Dublin), Bryan O'Mahony (Aontas na Mac Léinn in Éirinn), Nessa McEniff (Learnovate), Rosemary Day (Mary Immaculate College), Leo Casey (National College of Ireland), Alison Cook-Sather (Bryn Mawr College), Jan McArthur (Lancaster University), Paul McSweeney (University College Cork), Áine Ní Shé (Munster Technological University), Frances O'Connell (Technological University of the Shannon), Emma Muldoon Ryan (Aontas na Mac Léinn in Éirinn), Tim Thompson (Maynooth University), Garrett Murray (Enterprise Ireland), Niamh Kennedy (National Student Engagement Programme), Buster Whelan (Trinity College Dublin Students' Union), Eamon Costello (Dublin City University), Ann Riordan (University College Cork).

HEA Generative AI Policy Framework
https://hub.teachingandlearning.ie/genai/policy-framework

HEA Generative AI Resource Portal
https://hub.teachingandlearning.ie/genai/

Generative AI in Higher Education Teaching & Learning:
Principles for Ethical AI Adoption

Version 1.0, December 2025
DOI: 10.82110/qmt6-jw48

Higher Education Authority, Dublin

How to cite:
O'Sullivan, James, Colin Lowry, Ross Woods & Tim Conlon. *Generative AI in Higher Education Teaching & Learning: Principles for Ethical AI Adoption.* Higher Education Authority, 2025. DOI: 10.82110/qmt6-jw48.

This document, and all original content contained within, is licensed under the Creative Commons Attribution-ShareAlike 4.0 International Public License (CC BY-SA 4.0).

Table of Contents

1
Generative AI Adoption in Irish Higher Education

The Irish higher education sector has already taken important first steps toward framing the opportunities and challenges of generative AI (gen AI). The initial set of ten recommendations for gen AI adoption published by the Higher Education Authority (HEA) provided institutions with an accessible orientation for policy and practice.[1] Those principles captured the immediate questions facing the sector – ranging from AI literacy and academic integrity to issues of privacy, sustainability, and sovereignty – and opened a national dialogue on how best to respond to the challenges posed to education by gen AI. The following principles, which frame the national policy framework set out by this and accompanying documents, are a direct continuation of that work:

(1) Academic integrity, transparency, and accountability;

(2) Equity and inclusion;

(3) Critical engagement, human oversight, and AI literacy;

(4) Privacy and data governance;

(5) Sustainable pedagogy.

These principles establish a set of core ethical commitments that should guide all institutional decisions on AI adoption, providing a common baseline for Irish higher education institutes (HEIs) that is aligned with both European regulatory requirements and wider international standards of responsible AI use.

Taking such a values-led approach is essential because gen AI, particularly large language models (LLMs) like ChatGPT, radically reshapes how knowledge is produced and the validity with which it can be assessed. Decisions about whether to integrate gen AI into teaching and learning cannot be reduced to questions of efficiency or competitiveness. They must be framed in terms of educational purpose, what kind of learning experiences we want to foster, what skills and dispositions students should develop, and what forms of scholarly engagement we want to protect.

Values-based guidance ensures coherence across institutions. Without shared principles, adoption risks becoming fragmented, with some HEIs embracing permissive experimentation while others default to restrictive prohibition, creating confusion for students and undermining trust.

A principled stance ensures that generative AI is aligned with the mission of higher education as a public good, protecting human dignity, advancing equity of access, and safeguarding the intellectual integrity upon which democratic societies depend.

These provisions are therefore not abstract aspirations but operational standards. They are designed to anchor institutional choices in fairness, transparency, accountability, and human agency, while also addressing the sector-specific challenges of workload, assessment, and student engagement. They serve as an ethical compass for educators, ensuring that experimentation with AI remains accountable to the values that define higher education.

1 O'Sullivan and Lowry, 'Ten Considerations for Generative Artificial Intelligence Adoption in Irish Higher Education.'

By adopting this values-led framework, HEIs will be positioned to integrate generative AI in ways that strengthen, rather than erode, the trust placed in universities. The ethical adoption of generative AI must be approached with rigour and integrity. It is not sufficient for institutions to express a commitment to ethical principles while, in practice, implementing tools or practices that compromise those very values. Sustained alignment between stated values and institutional actions is essential to ensuring that the use of generative AI strengthens, rather than diminishes, the trust placed in higher education. An institution's approach to artificial intelligence can only be regarded as ethically robust if it demonstrates coherence between the systems it adopts and the values articulated in this framework. Where misalignment occurs, there is a risk of exposing students and staff to harm and of diminishing public confidence in higher education as a trusted steward of knowledge, integrity, and the public good.

The credibility of the higher education sector rests on the consistent alignment of institutional behaviour with declared values. This entails not only compliance with relevant regulations but also active discernment in determining which tools are adopted, how they are deployed, and whether their use genuinely advances the educational mission. Where tools cannot be demonstrated to meet these ethical standards, their use should be reconsidered, and where risks are identified, they should be managed transparently and with appropriate oversight.

These values are fundamental rather than discretionary, and they constitute the standards against which institutional decisions on generative AI should be measured. The provisions outlined in this framework are designed to ensure that higher education in Ireland engages with gen AI in ways that are not only innovative but also credible and trustworthy. They set the clear expectation that the integration of generative AI into teaching and learning will be conducted in a manner consistent with the highest ethical commitments of the sector.

Each of the five principles addresses a distinct domain of responsibility, but they function as interdependent dimensions of a single framework. No principle operates in isolation. Decisions made under one create obligations and constraints for the others, and their collective strength derives from mutual reinforcement rather than independent application.

Mapping these interdependencies shows how academic integrity relies on equity, literacy, privacy, and sustainability provisions, how equity cannot be separated from the other four domains, and how each principle shapes and is shaped by the others. Understanding these relationships is essential for coherent implementation, as institutions cannot pursue compliance in one area while neglecting its implications elsewhere.

Beyond the specific interconnections between individual principles, certain requirements and themes recur across the entire framework. These cross-cutting elements represent those recommendations that might be considered as the core foundations of ethical AI adoption and the recurring obligations that every implementation decision should satisfy:

(1) Human accountability cannot be delegated: Human judgement must be preserved in assessment, oversight, and all determinations relating to teaching and learning.

(2) Legal compliance is the minimum, not the goal: Compliance with GDPR and equality legislation establishes the floor, while ethical practice determines the ceiling.

(3) No AI tool is ethically complete: Openly acknowledge and document limitations and compromises and, where possible, implement compensating safeguards.

(4) Transparency is a foundational requirement: Make gen AI use, procurement criteria, data practices, and system operations visible to all stakeholders.

(5) Both adoption and refusal should be critical: Thoughtful rejection of gen AI tools is as valid as critical integration when pedagogical judgement demands it, but institutions should make clear to all staff that complete refusal is technically impracticable given AI's embedding in mainstream software and workflows.

(6) Support and enable staff: Institutions have a responsibility to equip staff with the knowledge and support required to engage with generative AI confidently and ethically. Professional development should be sustained, accessible across all roles, and integrated into existing frameworks for workload and recognition, regardless of whether individual staff intend to implement, investigate, or caution gen AI in teaching and learning.

(7) Discipline-specific AI literacy: Staff and students across all programmes should develop a critical understanding of generative AI, encompassing its technical foundations, disciplinary applications, ethical implications, and broader societal impacts. Literacy should not be optional or confined to technical fields.

(8) Students remain responsible for work they submit: AI assistance does not transfer accountability. Students answer for the accuracy and integrity of all submissions.

(9) Institutional infrastructure and resourcing: Institutions should maintain adequate infrastructure and capacity to support the ethical and effective use of generative AI, including appropriate licences, governance and compliance mechanisms, and monitoring systems.

(10) AI sovereignty and vendor independence: Institutions should safeguard their sovereignty over data, systems, and decision-making, and take active steps to avoid vendor dependencies that could compromise institutional autonomy.

(11) Embed equity in every decision: Equity is not a standalone concern but a dimension of every policy choice, procurement decision, and pedagogical practice.

(12) Procurement as ethical governance: Tool approval processes should verify compliance, assess bias, ensure equitable access, and evaluate environmental impact.

(13) Regular review and adaptation: Review regularly and retire tools when appropriate. Policies and practices should be updated in light of new evidence, monitored impacts, and ongoing technological developments.

(14) Evidence-based practice and continuous improvement: Base decisions on evidence rather than assumption. Institutions should commit to transparent reporting and demonstrate a readiness to revise or discontinue approaches that prove ineffective.

These requirements exist to serve a single overriding commitment: the adoption of generative AI in higher education must strengthen, not erode, the institution's capacity to fulfil its mission as guarantor of legitimate knowledge and contributor to the public good. Decisions about procurement, pedagogy, assessment, and governance should be tested against this standard. Where AI tools or practices cannot meet it, such as where they compromise integrity, deepen inequity, evade accountability, violate privacy, or undermine the intellectual capabilities that define educated persons, they should not be adopted, irrespective of convenience or competitive pressure.

2
Provisions of the Principles for Ethical AI Adoption

This document establishes the core foundations for the responsible introduction and sustained use of generative AI in higher education teaching and learning. Each provision addresses a distinct dimension of responsible adoption, including how we maintain academic standards, what transparency we owe to our communities, how we evaluate and procure tools, and where boundaries must be drawn around acceptable use. These provisions should be read as recommendations intended to shape procurement decisions, classroom practice, and institutional governance. They recognise that artificial intelligence in higher education presents risks that cannot be managed through individual discretion alone, requiring instead coordinated policy that protects students, preserves academic integrity, and ensures institutions remain accountable for the technologies they endorse.

2.1 Academic Integrity, Transparency, and Accountability

Academic integrity, transparency, and accountability form the cornerstone of this framework and address the epistemological challenges that generative AI poses to knowledge creation and validation. The institution's role as guardian of legitimate knowledge requires transparent disclosure of gen AI involvement in academic work, not just as procedural compliance but as recognition that the provenance of ideas matters to their evaluation and development. Accountability mechanisms ensure that the convenience of AI assistance does not erode the rigorous intellectual labour that transforms information into understanding. This principle acknowledges that while gen AI can augment human cognition, the responsibility for academic work must remain traceable to human agents who can defend, contextualise, and take ownership of intellectual contributions.

2.1.1 Institutional policies and academic freedom

Every HEI should publish a single, institution-wide policy on the pedagogically appropriate uses of gen AI in teaching and learning. Policies should set discipline-sensitive exemplars within a consistent core rule set that defines permitted, conditional, and prohibited uses for learning and assessment.

Only AI tools that have passed institutional procurement, data-protection, and ethics review should be mandated for student use. Reviews should document GDPR compliance, data minimisation, storage location, vendor security, ethical considerations, and model-use constraints, such as age limits, permitted purposes, data-handling rules, and restrictions on reuse of outputs. The results of these reviews, and the relevant constraints, should be communicated to staff and students in plain language so that no one is required to interpret complex licensing terms or hidden data practices on their own. Where a tool cannot meet institutional privacy and data-governance requirements, staff may not require its use and must offer an equivalent non-AI pathway without disadvantage.

Institutional policies should also acknowledge that no gen AI system satisfies all ethical requirements. Large language models carry, for example, unresolved risks of bias, opacity, environmental impact, and data leakage. Procurement and ethics reviews should therefore operate on a principle of risk acceptance with mitigation, clearly identifying the risks that cannot be eliminated, explaining why the tool is still being adopted, and setting compensating safeguards such as disclosure requirements and human oversight. Staff and students must be made aware that adoption represents a managed compromise rather than assurance of perfect ethical compliance.

HEIs should further establish a duty to review and, if necessary, retire AI tools where risks become unacceptable or where regulatory guidance changes. Reviews should be scheduled at least annually, with any material change to a vendor's terms, model behaviour, or data-handling practices triggering re-evaluation. Where risks cannot be mitigated, the tool should be withdrawn from mandated use, with prompt communication to staff and students and provision of alternative arrangements.

Transparency is of paramount importance in the procurement and ethics review processes. Each HEI should maintain and publish an institutional register of approved AI tools. For every approved tool, the register must provide staff and students with a clear summary of the criteria applied during evaluation, including GDPR compliance checks, data-governance standards, vendor security protocols, accessibility standards, and environmental impact considerations. The summary must detail the methods used to assess the tool, explaining how bias was tested, how privacy safeguards were verified, and what evidence or documentation the vendor provided to support their claims. The register should document both what the review found and what it could not resolve, identifying which risks were successfully mitigated while acknowledging any residual risks that remain. Each entry must include the decision rationale, explaining why the tool received approval despite unresolved risks and specifying what safeguards or conditions were imposed.

The register should outline the monitoring plan, detailing how the tool will be tracked over time and what circumstances would trigger a new review or retirement. The register must be updated whenever a tool is newly approved, re-evaluated, or retired, and should be accessible to all staff and students. Transparency in process is as important as transparency in use, as without visibility of how risks were assessed and mitigated, neither staff nor students can credibly uphold integrity and accountability in gen AI adoption.

Policies should establish a duty of transparency for both staff and students. Any substantive AI assistance in academic work should be disclosed using a standard AI use declaration that identifies the tool, the purpose of use, the nature and extent of contribution, and the author's verification steps.[2]

Policies should affirm academic freedom and pedagogical autonomy. Staff retain the discretion to choose appropriate methods and materials, consistent with institutional statutes and applicable law. This freedom, however, operates within binding requirements of integrity, transparency, privacy, and equity. Gen AI is now embedded in mainstream software, research workflows, and student study habits, meaning complete opt-out approaches are essentially impractical, and also risk inequity. European guidance emphasises human oversight, transparency, privacy, and accountability in educational AI use, while sector analyses show rapid uptake and recommend moving from reactive prohibition to governed enablement. Educators therefore have dual responsibilities to preserve standards through secure, transparent assessment and to prepare students for the realities of an AI-enabled society. Academic freedom is not freedom from these sectoral duties, but freedom in choosing the most effective ways to meet them.

For this duty to be workable, HEIs must provide structured and coordinated support. These responsibilities cannot rest solely with individual academics; they require institutional systems that provide consistent support for assessment design, compliance, and AI literacy. Institutions should therefore develop coherent systems that

[2] For illustratives example of AI use declarations, see relevant academic style guides.

provide professional development through mandatory and ongoing training on assessment design in the age of AI, the use of disclosure statements, and the limits of detection technologies. They must establish shared resources to maintain repositories of sample briefs, disclosure exemplars, and discipline-specific case studies to reduce duplication of effort. Technical assurance should come through a central list of institutionally approved tools with clear data-protection reviews, so that individual staff are not forced to make compliance judgements alone. Advisory support through academic integrity and AI advisory groups at school or faculty level enables staff to seek guidance on difficult cases without fear of inconsistency or reputational risk. Institutions must recognise the additional workload of assessment redesign and integrity management in workload models, promotions, and teaching awards.

Without these supports, the obligation to integrate AI responsibly risks falling unevenly on individual staff members, creating confusion and inconsistency across programmes. Institutional leadership must therefore take responsibility for providing the frameworks, training, and assurance that allow academic freedom to flourish within a clearly governed environment.

2.1.2 Assessment design and authenticity

Assessment design in the age of generative AI must contend with a set of interlinked challenges that pose a significant threat to academic integrity and learning assurance, including the need to assure validity when gen AI can produce passable work and detectors are unreliable. Authorship should remain visible through diversified, triangulated tasks that evidence process, judgement, and feedback while remaining inclusive at programme level. Staff should operationalise clear AI permissions, disclosures, and marking implications inside briefs without creating ambiguous boundaries or over-relying on detection. They should balance security with feasibility by adding observed or dialogic elements and novel tasks without unsuitable workload, and they must meet equity, privacy, and tool-governance duties when AI is permitted, including approved-tool use, non-AI alternatives, and explicit access provisions.

These pressures mean assessment design is no longer a purely pedagogical exercise, but a cross-cutting policy concern that requires clarity of purpose, institutional support, and regular review to keep pace with changing AI capabilities and regulatory expectations.

Recognising discipline-specific challenges is essential for credible assessment reform. Generative AI does not affect all fields in the same way. Its capacity to generate fluent text may undermine traditional essays in the humanities, while in design or computer science AI-supported iteration may itself become a legitimate object of learning. In clinical or professional programmes, however, unaided performance, ethical reasoning, and real-time judgement remain non-negotiable, making heavy reliance on AI inauthentic or unsafe. Language learning presents its own difficulties, since gen AI can easily handle tasks intended to build vocabulary and syntax through active production. These differences mean that a single institutional rule or generic framework, if applied rigidly, risks either over-restricting innovation in one discipline while eroding integrity in another. Policy should therefore mandate a consistent baseline while allowing individual units, programme teams, and teaching staff to adapt assessment design, exemplars, and marking criteria to the epistemic norms and professional standards of their fields and coursework. Without this recognition, institutions risk undermining both pedagogical credibility and external trust in their awards.

Discipline-specific challenges cannot be addressed unless staff have a baseline of AI literacy that enables them to understand both the capabilities and the limits of the tools. Generic warnings or broad institutional rules are insufficient, as subject specialists must be able to interrogate what AI means in the context of their own field. For example, a historian needs to understand how training data biases may distort narratives just as much as a computer scientist should be able to evaluate code generated by large language models for efficiency and security flaws. In design or creative arts, critical literacy is required to distinguish between AI-generated artefacts and authentic practice, and to ensure students can justify aesthetic or conceptual choices rather than outsourcing creativity. Professional fields such as medicine, law, or teacher education demand even sharper literacy, as practitioners must be trained to question the factual accuracy, ethical compliance, and professional standards of AI-supported outputs.

Embedding AI literacy training across specific disciplines, as opposed to institutions, allows subject specialists to take ownership of these questions. It equips staff to design assessments that target authentic disciplinary learning outcomes rather than generic skills, and it helps students to situate gen AI within their own professional trajectory. Without such literacy, staff risk applying either blanket prohibitions or uncritical permissions, and students risk substituting tool use for genuine disciplinary competence. Recognising discipline-specific challenges and equipping both educators and learners with the literacy to navigate them is therefore critical to ensuring fairness and preparing graduates for responsible participation in their respective professions.

But there are general best practices which should be adopted on a transdisciplinary basis. Every assessment brief should contain an AI-use statement that declares whether AI is prohibited or permitted for co-creation with critical evaluation. The statement should specify disclosure and citation duties, and link the permitted AI use to marking criteria.

Each statement should name permissible tools or classes of tools, describe permitted purposes, set boundaries on what must be the student's own work, and indeed, what constitutes 'own', set disclosure format and location, and explain assessment consequences for misuse. Where AI use is permitted or encouraged, the brief must confirm that approved access is available or that a non-AI path exists without disadvantage.

Permissions should be expressed in plain language using bands that map to an intuitive scale of use. Popular frameworks like the Artificial Intelligence Assessment Scale (AIAS)[3] are a useful signalling device that allow academics and students to talk plainly about what kinds of AI use are acceptable in a given task. They do not, by themselves, secure validity, fairness, or privacy. Institutions should adopt such frameworks with governance, equity, and assessment design measures that address risks that scales like AIAS leave open.

The AIAS, a particularly popular framework, provides clarity, but boundaries can be ambiguous in practice. Terms like 'surface-level editing' versus 'substantive content' invite divergent interpretations unless tightly exemplified in rubrics. Higher bands increase administrative load, demanding prompt logs or draft histories that

3 Perkins et al., 'The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment.'

can shift effort from learning to documentation. There are equity risks here, as students with premium features or better devices gain advantages unless institutions provide access or alternatives. Privacy and governance requirements are not embedded in the scale, yet EU and Irish guidance make data protection, transparency, and human oversight first-order duties. At higher bands students may optimise for prompt craft rather than disciplinary mastery, unless rubrics re-centre intended outcomes and keep the teacher firmly 'in the loop'. The scale travels well into design, coding, and studio contexts but poorly into disciplines where unaided performance is essential, such as clinical judgement or initial language acquisition. Furthermore, model capabilities evolve quickly, and what counted as 'limited support' yesterday may now amount to substantive authorship, requiring institutions to version their implementations, schedule annual reviews, and update exemplars accordingly.

For these reasons, HEIs should support a systematic renewal of assessment design to reduce substitution risks while preserving validity and equity. Programme teams should prefer tasks that make authorship and judgment visible through activities such as staged drafts with feedback, unique datasets, supervised artefact creation, oral explanations, code walkthroughs, lab or studio notebooks, or viva voce.[4] Where traditional written tasks are retained, briefs should incorporate situated or reflective elements that require human reasoning tied to taught materials. Only by combining baseline institutional rules with discipline-sensitive literacy and design can higher education maintain integrity and fairness in the age of generative AI.

Designing new forms of authentic or process-based assessment requires time, training, and, in some cases, additional infrastructure. Without structured institutional support, these demands risk being absorbed into current workload pressures, with potential consequences for both quality and sustainability. Policy should therefore ensure that changes to assessment design and delivery are recognised within workload models, appropriately costed in programme planning, and supported through professional development and technical assistance. AI-resilient assessment requires collective commitment and planning to sustain staff capacity and protect student learning outcomes.

2.1.3 Disclosure, citation, and authorship

The integration of generative AI into academic work necessitates a fundamental reconceptualisation of authorship, attribution, and intellectual responsibility. Where traditional academic integrity frameworks assumed human-only production, the advent of gen AI creates new categories of contribution that must be made visible and assessable. HEIs should establish disclosure and citation requirements that preserve the chain of intellectual accountability while enabling legitimate AI-supported learning.

Where AI is used within permitted bounds, students should include a comprehensive AI use declaration at the point of submission.[5] Any such declaration is a substantive component of academic integrity that enables assessors to evaluate the authenticity and quality of student work. The declaration must be specific enough to support verification and must identify the exact name and version of the AI system used, including any plugins,

4 For a discussion of AI-resilient assessment practices, see supporting instruments.

5 See established style guides on AI citation practices, which continue to develop.

extensions, or additional features employed. Students must provide a clear description of how the AI was used, whether for ideation, research synthesis, code generation, language editing, or visual creation. The declaration should include representative prompts or describe the general prompting strategy employed, sufficient for assessors to understand the nature of human-AI interaction.

Crucial to the declaration is a detailed account of the verification process. Students should specify the steps taken to validate gen AI outputs, including fact-checking methods and corrections made. They must also provide an honest evaluation of the relative contributions of human and AI to the final work, expressed in terms that map directly to the assessment criteria. This transparency allows assessors to calibrate their evaluation appropriately and ensures students remain conscious of their own intellectual contribution throughout the process.

These declarations should be incorporated into the work itself, not relegated to separate documents that may become detached. For written assignments, the declaration should appear immediately after the title page or in a designated section of the methodology. For code submissions, it should be included in documentation headers or README files. For creative works, it should form part of the artist's statement or design rationale. This integration ensures the declaration travels with the work through all stages of assessment and review.

AI-generated content must be cited according to evolving scholarly conventions that recognise both the tool and the human operator as agents in the creative process.[6] HEIs should adopt and maintain citation guidelines that distinguish between levels of assistance, from surface-level editing to substantial content generation to collaborative iteration. These distinctions matter because they signal different degrees of intellectual contribution and allow assessors to evaluate work appropriately.

Where AI summarises or paraphrases existing sources, those original sources must still be cited directly, not laundered through AI attribution. This dual citation requirement ensures the intellectual genealogy of ideas remains traceable and that original authors receive proper credit. For work involving multiple rounds of human-AI interaction, citation should capture the evolutionary nature of the collaboration, documenting how ideas developed through iterative refinement. Where relevant to reproducibility or assessment, citations should include model parameters, temperature settings, or other technical specifications that shaped the output.

Citation formats must evolve with disciplinary norms. While humanities disciplines may emphasise narrative descriptions of AI use that situate the technology within broader methodological frameworks, STEM fields may require more technical documentation including version control logs, prompt repositories, or computational notebooks that enable reproducibility. Professional programmes must align citation practices with sector standards, ensuring students develop habits transferable to practice. This disciplinary sensitivity prevents citation requirements from becoming either meaninglessly generic or inappropriately prescriptive.

Students are the authors of work they submit and bear full responsibility for its content, accuracy, and integrity. This fundamental principle does not change when AI tools are used within permitted bounds. Students are obligated to verify all AI-generated claims against authoritative sources, with particular attention to statistical data, historical events, scientific facts, and citations that AI systems are known to fabricate. They must identify and

6 It is advised that teaching staff keep up-to-date with established academic style guides on matters relating to gen AI citation practices, which continue to develop.

address biases in gen AI outputs, including stereotypes, cultural misrepresentations, and discriminatory framings that may reflect the limitations of training data. Students must be aware, and respond appropriately, to the reality that many gen AI models have been trained on and can reproduce copyrighted material without attribution. Most critically, they must demonstrate understanding of all submitted content through the ability to explain, defend, and elaborate on any aspect of the work. This means that errors, inaccuracies, or misrepresentations in AI-generated content will be assessed as if they were the student's own mistakes, maintaining the essential link between submission and accountability.

This responsibility extends beyond fact checking and includes critical engagement with AI outputs. Students must be able to justify why they accepted, modified, or rejected AI suggestions, demonstrating the kind of intellectual autonomy that distinguishes human learning from mechanical reproduction. The capacity to evaluate AI outputs critically, to recognise their limitations, and to improve upon their outputs represents a new form of academic literacy that should be developed and assessed across disciplines.

Academic staff should model the same disclosure and citation standards in all teaching materials, research outputs, and administrative documents. When lecture slides, handouts, reading lists, or sample solutions incorporate AI assistance, this must be clearly attributed. Even in routine communications such as email responses or discussion board posts, substantive use of AI assistance should be acknowledged.

In research supervision, supervisors must declare any AI use in reviewing student work or generating supervisory feedback. This transparency normalises disclosure practices and maintains the mutual accountability that underpins academic relationships. Staff cannot credibly enforce standards they do not themselves observe, and their modelling of good practice provides students with concrete examples of how to integrate AI responsibly into academic work.

To ensure disclosure and citation requirements are meaningful rather than performative, institutions must establish verification mechanisms that balance rigour with practicality. Quality units should periodically sample submitted work to verify that AI disclosures accurately reflect the level of assistance received. Where institutions use metadata, version histories, or interaction logs to corroborate disclosure statements, such technical verification must comply with data-protection requirements and must not serve as the sole basis for integrity determinations.

Students must be aware that they may be required to explain their AI use in viva voce examinations or portfolio defences, demonstrating understanding of both the tool's contribution and their own intellectual process.

2.1.4 Detection and investigation

AI-related misconduct differs from traditional plagiarism. Whereas plagiarism involves the unacknowledged reuse of existing material, generative AI can create novel outputs that nonetheless breach traditional notions of authorship and intellectual property. Institutions must therefore re-examine investigative practices with these new conditions in mind.

AI detection tools cannot provide a reliable solution. AI-use indicators and content-classification systems produce probabilistic scores rather than definitive findings, are vulnerable to evasion, and are prone to both false positives and false negatives.[7] These systems typically rely on proxies like perplexity and burstiness, metrics of predictability and sentence variation, which may have been marginally useful against earlier models but fail against contemporary systems that convincingly mimic human rhythm and style. The consequence is human work written in more formal or academic registers is disproportionately flagged as AI, while AI-generated text prompted for casual tone often escapes notice. Students who write in formal or conventionalised styles, particularly non-native speakers, are disproportionately flagged,[8] while sophisticated gen AI users may escape detection through prompt engineering or hybridising AI and human work. Tools trained on older models degrade quickly as new systems emerge, creating an unwinnable technological race.

Beyond the statistical and technical limitations, detection tools operate on assumptions about what 'real' writing should look like, treating fluency, coherence, and formal register as potential signs of artificiality. By embedding narrow norms of natural writing, these tools reproduce structural inequities, rewarding those whose style falls outside the training-data uncanny valley and punishing those whose work most closely resembles the genres that models were trained on. The result is a perverse inversion of academic values, wherein clarity and control are treated as suspect, while mediocrity or idiosyncrasy passes unnoticed. This reveals why detection is not merely unreliable but actively harmful: it shifts the burden of proof onto the student, undermines trust in academic judgment, and reinscribes linguistic and cultural bias under the guise of neutrality.

For these reasons, AI detection systems are not recommended as primary evidence in higher education, and serve no role as sole or determinative proof of misconduct.

Investigations must rest on principles of natural justice and procedural fairness. Students must be afforded the presumption of innocence, with the burden of proof borne by the institution. Findings must be reached on the balance of probabilities, using clear, documented evidence. Responses must be proportionate to the seriousness of the allegation: suspected misuse in a minor formative task cannot be treated in the same way as substitution in a capstone or professional placement. Timeliness is also essential: cases must be progressed promptly to minimise stress and preserve learning opportunities, while allowing sufficient time for a thorough review.

Evidence must be triangulated. Institutions should give primary weight to process evidence, such as draft histories, version logs, and research trails that show natural development of work. Comparative analysis across a student's portfolio may provide useful context. Technical indicators such as metadata or citation anomalies may be considered, but only as supplementary signals. Circumstantial evidence, such as the absence of drafts, must be weighed carefully, since legitimate alternative explanations are always possible.

A critical safeguard is the oral or live assessment. Where a staff member has credible grounds to believe that a submitted assessment does not represent the student's own work, the student should be offered the opportunity to demonstrate authorship directly. Institutions should develop a policy provision that allows for any student across all modules to be called, at the request of the module coordinator, to an oral examination,

7 Otterbacher, 'Why Technical Solutions for Detecting AI-Generated Content in Research and Education Are Insufficient.'

8 Liang et al., 'GPT Detectors Are Biased against Non-Native English Writers.'

viva, code walk-through, studio critique, or equivalent dialogic exercise in which they account for and extend their submitted work. The policy should clearly state that, in all instances, oral assessment overrides the written artefact: if the student can demonstrate understanding, reasoning, and command of sources or methods, the oral performance confirms authenticity and secures credit. Conversely, inability to explain or extend the work constitutes strong evidence of inauthentic authorship and may warrant sanction.

To ensure fairness and consistency, every such examination may involve the staff member who made the referral, but also, a panel that includes other academic staff with relevant expertise, at least one of whom must come from outside the examining unit, so that decisions are not made unilaterally and students are protected from arbitrary referral or judgement. Institutional policies of this nature do not preclude staff from having their own, module-level oral and live examination processes, but equally, the presence of such does not negate any institutional provisions.

Investigations and oral examinations must be conducted fairly and transparently. Students must be informed in writing of the precise concerns, the evidence under review, their rights, and the supports available. Decisions must be communicated in writing, with reasons clearly explained and rights of appeal available. Appeals must be heard by independent staff not involved in the original decision and resolved within defined timeframes to protect student progression.

All staff involved should have the opportunity to be trained in the limits of detection software, evidentiary standards, interviewing techniques that are rigorous but non-intimidating, unconscious bias awareness, GDPR compliance, and recognising when education rather than punishment is the appropriate outcome. Training should be refreshed regularly, and supported by practical case studies and shared sectoral exemplars to ensure consistency across faculties. Institutions should ensure that investigation processes are adequately supported through appropriate time allocations within workload models, access to administrative assistance, and explicit recognition within institutional planning and professional development frameworks.

The purpose of investigation is not only enforcement but the preservation of trust in assessment and the protection of fairness. Sanctions have a place where deliberate deception is proven, but the primary goal remains developmental, in helping students understand expectations, protecting those who use AI responsibly and transparently, and ensuring that Irish higher education awards remain credible and trusted.

2.1.5 Capacity building

The integration of gen AI into higher education depends on the capacity of staff and students to engage with these tools critically, ethically, and effectively. This capacity must not be assumed and requires systematic and sustained institutional support.

HEIs should implement mandatory staff development programmes addressing both technical understanding and pedagogical practice. All teaching staff must complete core training that provides a conceptual grasp of how generative AI works, including how training data shapes outputs and embeds bias, the capabilities and limitations of current systems, and the distinctions between text, image, code, and multimodal models. Training should also cover the ethical implications of bias, opacity, and environmental impact.

Assessment design represents the most urgent training need, and staff must be able to design authentic assessments that resist substitution while remaining inclusive. This includes learning to balance security and pedagogical goals, to construct rubrics that account for disclosed AI assistance, and to write clear AI-use statements for briefs. Staff should understand the workload implications of such assessments and plan accordingly, with workload models recognising the additional time required.

Capacity building must also extend to integrity procedures. Staff involved in misconduct investigations should be trained in evidentiary standards, the limitations of detection tools, the triangulation of multiple evidence sources, and fair interviewing techniques. They must also understand GDPR and intellectual-property obligations, and be able to distinguish between deliberate misconduct and misunderstandings.

Student induction at all levels should include structured AI literacy. Students must develop a conceptual understanding of generative AI, awareness of system limits, and skills in verifying and critiquing outputs. They must also be taught disclosure and citation practices, privacy and security responsibilities, and discipline-specific standards for ethical use. AI literacy should be embedded across curricula through integrated assignments, reflective portfolios, and case studies, not treated as a standalone module.

HEIs should maintain the systems and structures necessary to support sustained capacity-building. This may include designated coordination for AI education, shared repositories of exemplars and guidance, sandbox environments for safe experimentation, and accessible support for both technical and pedagogical queries. Related content should be reviewed regularly to reflect technological developments, regulatory updates, and sectoral feedback.

Adequate institutional provision is essential. Workload planning, equitable tool access, and coordinated development of expertise and systems are necessary to meet policy objectives.

2.1.6 Governance, monitoring, and review

Effective governance of AI in higher education requires clear academic ownership, systematic monitoring, and responsive review mechanisms.

Ultimate authority over institutional AI policy in teaching and learning should rest with the academic council or equivalent body, which should adopt institutional policies, approve major revisions, and ensure adoption supports institutional mission and strategy while keeping pedagogical decisions grounded in academic judgement. Governance bodies should receive regular reports on risk assessment covering academic, reputational, and operational dimensions, and should advise on resource priorities to ensure resources align with academic aims.

Academic units are responsible for contextualised implementation, adapting institutional policy to disciplinary contexts, ensuring all assessment briefs include clear AI statements, and providing discipline-specific training and development. Units should also coordinate programme-level AI literacy to ensure progression and coherence, while leading innovation through pilot projects in AI-enhanced pedagogy.

Central support services must provide the assurance and infrastructure necessary for consistent implementation, maintaining a register of approved tools, undertaking institutional procurement and security reviews, managing licencing and equitable access, providing staff and student training, and conducting regular audits of compliance that include sampling assessment briefs for compliance and reviewing integrity cases for fairness and consistency.

Institutions should introduce proportionate compliance monitoring to track implementation and support continuous improvement. Indicators might include the proportion of modules containing compliant AI statements, participation rates in staff training, and the timeliness of integrity case resolution. These should be complemented by quality measures that assess effectiveness – such as clarity of policy communication, staff confidence, and evidence of pedagogical innovation. Continuous monitoring of risk indicators should include the proportion of AI-related integrity cases, appeals, data protection incidents, and vendor compliance issues.

Regular reporting contributes to transparency and accountability. Institutions are encouraged to provide an annual summary to academic governance bodies outlining compliance, quality, risk, and impact.

Annual review should incorporate monitoring evidence, technological advances, regulatory updates, and stakeholder consultation, ensuring approved tools continue to meet instructional requirements with retirement procedures invoked where necessary. Continuous updates to guidance and resources should reflect current practice, with trigger-based reviews initiated in response to major capability breakthroughs, regulatory changes, significant incidents, or shifts in institutional strategy.

HEIs must ensure transparency through publicly accessible AI policies, a maintained register of approved tools, and publication of summary data on use and impact. Institutions should also establish clear contact points for queries and complaints, and maintain open channels for communicating policy updates. Accountability structures should assign responsibilities unambiguously, specify decision rights, and maintain clear and transparent escalation procedures.

2.1.7 Summary of Recommendations

(1) Publish a single institutional AI policy setting permitted and prohibited uses across teaching and assessment, with discipline-sensitive exemplars and protection of academic freedom.

(2) Mandate institutional approval for tools so that only AI systems that pass procurement, GDPR/data-protection, and transparent, documented ethics reviews may be required for student use.

(3) Maintain a public register of approved tools, updated regularly with review criteria, risks, safeguards, monitoring plans, and retirement decisions.

(4) Provide institutional supports for professional development, repositories of exemplars, advisory services, equitable access to approved tools, and workload recognition.

(5) Require disclosure of AI use through a standard declaration specifying tools, purpose, extent, and verification of outputs.

(6) Mandate citation of AI outputs and sources, distinguishing levels of assistance and aligning with disciplinary conventions.

(7) Preserve student accountability so students remain fully responsible for accuracy and integrity of submitted work, and staff must model disclosure in teaching and supervision.

(8) Embed AI-use statements in every assessment brief, specifying permissions, disclosure duties, marking implications, and, where necessary, alternative non-AI pathways.

(9) Adapt and further develop validated frameworks (eg. AI Assessment Scale)[9] to signal allowable AI use, supplemented by institutional rules on gen AI governance.

(10) Redesign assessments for authenticity, favouring approaches that make authorship and judgment visible.

(11) Resource assessment reform through workload allocation and programme planning supports.

(12) Ensure disciplinary sensitivity, supporting programme teams to adapt assessment rules to field-specific norms while upholding baseline integrity and equity standards.

(13) Integrate AI literacy across programmes, enabling staff and students to critically evaluate AI in the context of their specific disciplines.

(14) Prohibit the use of AI detectors and probabilistic tools as determinative evidence of misconduct.

(15) Ground integrity investigations in natural justice, including presumption of innocence, balance of probabilities, proportionate sanctions, and timely resolution.

(16) Triangulate evidence in investigations, using drafts, process records, and oral demonstrations.

(17) Provide an institution-wide oral assessment safeguard wherein students must be able to demonstrate authorship live, with oral performance overriding written artefacts.

(18) Ensure fairness and transparency in investigations, including written notice of concerns, access to evidence, rights of appeal, and involvement of independent panel members.

(19) Train and resource investigators in evidentiary standards, unconscious bias, GDPR, interviewing, and workload recognition.

(20) Establish governance, monitoring, and review, ensuring academic councils hold ultimate authority and institutions publish annual reports, conduct annual and trigger-based reviews, implement clear accountability lines, and resource governance functions adequately.

9 Perkins et al., 'The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment.'

2.2 Equity & Inclusion

The OECD warns that the unchecked spread of AI in education risks widening existing disparities.[10] Access to AI tools risks creating new forms of educational stratification that could exacerbate inequalities within society. This principle requires that HEIs actively mitigate differential access to AI technologies and ensure that their implementation does not privilege certain learning approaches, linguistic backgrounds, or socioeconomic positions. Equity goes beyond access, requiring that AI systems be scrutinised for embedded biases that might perpetuate historical disadvantages, particularly for communities traditionally underrepresented in higher education.

2.2.1 Commitment to inclusive education

The adoption of AI in higher education must be aligned with Ireland's obligations under equality legislation, the Public Sector Equality and Human Rights Duty,[11] and the national commitment to inclusive education under the UN Sustainable Development Goal 4, to 'ensure inclusive and equitable quality education and promote lifelong learning opportunities for all'. International guidance reinforces this, such as UNESCO's global frameworks on AI in education which stress that AI adoption must be grounded in humanistic values, including inclusion, equity, gender equality, and respect for cultural and linguistic diversity.[12]

Every HEI should include within its institutional AI policy a clear statement affirming equity and inclusion as guiding principles. This commitment should be accompanied by a transparent account of how it will be implemented through procurement processes, staff and student development, and curriculum design, ensuring that no cohort is disadvantaged in an AI-enabled learning environment. The commitment should also acknowledge intersectional disadvantage, recognising that students experiencing multiple forms of marginalisation face compounded rather than additive barriers.

2.2.2 Equitable access to AI tools and infrastructure

Equitable access to AI is not guaranteed. Many gen AI tools require reliable broadband and paid subscriptions. Students from lower socio-economic backgrounds or from regions with limited connectivity are at risk of being excluded from AI-enabled learning, so institutions should take proactive steps to ensure that engagement with AI does not depend on students' private means. This includes providing institutional licences for approved tools and ensuring that campuses are equipped with the hardware, software, and connectivity needed to support inclusive access, alongside mechanisms to provide individual access where necessary.

The risk of digital poverty extends beyond hardware and subscriptions. Students experiencing housing instability may lack quiet spaces for AI-assisted study. Working students may have limited time to develop AI literacy skills. Commuter students may struggle with campus-based AI resources. Institutions must recognise

---

10 Varsik and Vosberg, 'The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education.'

11 'Public Sector Equality and Human Rights Duty.'

12 Miao and Holmes, 'Guidance for Generative AI in Education and Research.'

these broader dimensions of digital exclusion in support strategies that address the full spectrum of access barriers. Possible measures include extending library access with dedicated AI-enabled study spaces, offering asynchronous AI-literacy training that accommodates work schedules, and ensuring that AI tools are mobile-optimised for students who rely primarily on smartphones.

2.2.3 Assessment equity and standardisation

The use of AI in assessment contexts presents particular equity challenges. Students with greater financial means might access more sophisticated AI tools for take-home assignments, creating unfair advantages over peers relying on free or institutional versions. Institutions should be alert to this emerging digital divide in assessment and take steps to preserve fairness and integrity.

Institutions should establish clear protocols for the use of AI in assessment that take account of differential access among students. Where AI use is permitted, HEIs may either designate specific, institutionally provided tools or require students to submit detailed declarations outlining which AI tools were used and for what purposes. Assessment rubrics should be adapted to evaluate critical engagement with AI rather than simply the sophistication of AI-generated content. For time-bound assessments, institutions might consider providing standardised AI access through controlled environments, ensuring all students work with the same tools under the same conditions.

Alternative assessment strategies that minimise advantage from differential AI access should be prioritised. These might include in-person presentations, reflective portfolios that document learning processes, collaborative projects where AI use is transparent and shared, or hybrid assessments combining AI-assisted preparation with non-AI demonstration of understanding.

2.2.4 Mitigating bias and discrimination in AI systems

AI systems are not neutral. They are trained on large datasets that frequently contain social biases related to race, gender, class, language, and disability. These biases, if left unaddressed, can be reproduced or amplified in outputs. The OECD has highlighted that unchecked adoption of AI can entrench inequities and undermine cultural responsiveness,[13] while the European Commission's guidelines for trustworthy AI explicitly identify diversity, non-discrimination, and fairness as key requirements for responsible practice.[14]

HEIs have a duty to ensure that the AI systems they approve for teaching and learning are subject to rigorous scrutiny for bias and discrimination. It is not sufficient for institutions to rely on vendor assurances or generic claims of compliance. Instead, this obligation must be operationalised through procurement and approval processes that apply explicit equity criteria to every AI system under consideration. These processes should require clear evidence of transparency in the provenance of training data, enabling institutions to assess

13 Varsik and Vosberg, 'The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education.'

14 'Ethics Guidelines for Trustworthy AI.'

whether datasets are representative, inclusive, and free from systemic patterns of exclusion. They should also require clarity regarding the system's intended educational use, specifying the contexts in which it can be relied upon and those where risks of biased or distorted outputs are greater. Where transparency cannot be demonstrated, institutions should take a precautionary approach, deferring adoption until sufficient evidence of equity and reliability is available.

A practical, equity-focused approval pathway should begin with an institutional review of any AI system proposed for use in teaching or assessment. This review should combine technical evaluation with ethical scrutiny. While full disclosure of training datasets or proprietary architectures may not be possible, vendors should be obliged to provide high-level documentation that enables institutions to make informed judgments. At a minimum, this should include statements of data governance practices, descriptions of representativeness efforts, and a clear outline of known limitations, risks, and inappropriate use cases.

Vendors should be required to specify the contexts in which their AI system is designed to be used in education, and the contexts in which it is not reliable. Reliability may be compromised in high-stakes assessment settings, in disciplines where factual precision is critical, or in applications involving vulnerable learners. By requiring vendors to delineate both the fit-for-purpose and not-fit-for-purpose use cases of their systems, institutions can better align tools with pedagogical objectives and protect students from inappropriate or harmful deployment.

Prior to full deployment, institutions should pilot AI tools in controlled settings, reviewing performance across diverse learner groups. Feedback from staff and students, including those from under-represented backgrounds, should inform decisions on whether the tool is approved for broader use and under what conditions. Review outcomes should be documented and registered transparently, with approvals revisited at regular intervals to ensure relevance as models and datasets change.

Beyond technical safeguards, staff and students must be supported to critically evaluate AI outputs, to recognise when bias is present, and to understand how it shapes knowledge production and representation. Bias in AI determines whose voices are included in teaching contexts, whose perspectives are excluded, and how students understand their own identity and belonging within higher education.

2.2.5 Irish language and minority language contexts

The Irish language performance of AI systems presents specific challenges for equity in higher education. Most generative AI tools demonstrate significantly reduced accuracy and fluency in Irish compared to English, potentially disadvantaging students in Irish-medium programmes and courses where bilingual competency is required. This linguistic inequity threatens commitments to Irish language education and cultural preservation.

Institutions should evaluate AI tools for Irish language capability as part of approval processes. Where tools demonstrate inadequate Irish language performance, compensatory measures should be evaluated. This might include dedicated support for Irish-medium students, alternative assessment arrangements that do not rely on AI assistance, or developing Irish language AI capabilities.

The higher education sector should take a coordinated approach to strengthening Irish-language capacity within AI systems. The sector should work collectively to promote improved Irish-language support in commercial tools and explore opportunities to support the development of solutions that advance equitable access to AI tools that adequately serve Irish language education.

2.2.6 Supporting diverse learners and needs

Generative AI has the potential to expand access to learning for students who face barriers in traditional modes of teaching. Research suggests that generative AI can scale personalised tutoring and help close quality gaps in teaching provision,[15] while international frameworks emphasise its capacity to provide translations, generate plain-language summaries, and enable multimodal functions such as text-to-speech or speech-to-text.[16] A systematic review of gen AI in special education further concluded that such tools can enhance personalised learning and social engagement for students with special needs, while highlighting ethical risks and uneven study quality.[17]

For students with disabilities, neurodiverse learners, and those studying in an additional language, these affordances can lower structural barriers and enable fuller participation in higher education. For second-language learners, gen AI can scaffold writing and comprehension through feedback and adaptive prompting strategies, yet these tools can also reinforce inequities by advantaging already digitally confident students and promoting over-imitation.[18]

Caution is essential, and the OECD warns against 'techno-ableism', the belief that technology alone can 'fix' disability, arguing that such assumptions can undermine systemic commitments to universal design and adequate student supports.[19] Claims that generative AI supports learning styles should also be avoided, as decades of educational research show no empirical benefit from tailoring teaching to self-reported learning styles. What is supported, however, is the principle of multiple representations, providing learners with different ways of accessing and engaging with material, consistent with Universal Design for Learning.

International students face additional challenges, as AI systems trained predominantly on Western educational contexts may produce outputs misaligned with their educational backgrounds or that perpetuate cultural biases. Institutions must ensure that AI implementation is culturally responsive, providing guidance on how AI outputs might reflect particular cultural assumptions and supporting international students in critically evaluating these limitations.

HEIs should position gen AI as one tool within a wider strategy for inclusive education. Students, particularly those from under-represented or marginalised groups, must be actively consulted about how these technologies affect their learning experience. Accessibility gains must be integrated into institutional commitments to universal design, inclusive pedagogy, and adequate support services.

15 Kestin et al., 'AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting.'

16 Elhussein et al., 'Shaping the Future of Learning: The Role of AI in Education 4.0.'

17 Voultsiou and Moussiades, 'A Systematic Review of AI, VR, and LLM Applications in Special Education.'

18 Warschauer et al., 'The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language.'

19 Varsik and Vosberg, 'The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education.'

2.2.7 AI literacy and professional development for equity

Without targeted interventions, those with existing technological expertise will benefit disproportionately, leaving behind students and staff who lack digital confidence or who face barriers to engaging with new tools. Both UNESCO[20] and the OECD[21] underline that equitable access to AI literacy training is essential if adoption is to be fair.

Equity considerations extend to staff. Sessional lecturers, teaching assistants, and staff on short-term contracts may have limited access to institutional AI training, professional development time, or participation in decisions about AI adoption that affect their teaching. HEIs should take steps to ensure that opportunities for AI-related training and engagement are available equitably across all staff groups, regardless of contract type or duration.

HEIs should seek to ensure that professional-development opportunities in AI literacy are accessible to all staff, across disciplines and employment types. Institutions should recognise the time required for such training within existing workload and development frameworks and schedule provision to accommodate part-time and multi-institutional staff. Where staff are required to adapt courses or design AI-related assessments, this work should be appropriately recognised within institutional planning and professional-development structures. Engagement with staff representative bodies can help ensure that these arrangements are equitable and responsive to the needs of all members of the academic community.

Professional development must address not only technical skills but also critical AI literacy, including recognising bias, understanding limitations, and making pedagogical decisions about appropriate use. Staff from disciplines less familiar with technology should receive additional support to ensure they are not disadvantaged in an AI-enabled teaching environment.

Students in every programme, not only those in technical fields, must be given the knowledge and critical skills needed to engage meaningfully with AI in their studies and future work.

AI literacy must encompass both functional competence and critical awareness, including the capacity to identify bias, evaluate limitations, and make ethical judgments about use. AI literacy programmes should be designed with accessibility in mind, providing multiple formats (video, text, interactive), flexible pacing, and support in multiple languages where student demographics warrant. Peer learning approaches that pair digitally confident students with those requiring support can build community while developing skills. Recognition may be given to students who complete AI literacy training through digital badges or micro-credentials that they can include in their professional portfolios.

Embedding these literacies across curricula and professional-development frameworks is essential to avoid a divided landscape between those able to engage meaningfully with AI and those excluded from its benefits.

---

20 Miao and Mutlu, 'AI Competency Framework for Teachers.'

21 'Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (Review Draft).'

2.2.8 Monitoring, accountability, and redress

Equity commitments require proportionate systems for monitoring and accountability. Institutions should integrate equity review of AI tool usage, examining access patterns, user satisfaction, and learning outcomes disaggregated by relevant demographic categories. These reviews should help identify whether AI adoption is narrowing or widening existing achievement gaps and inform ongoing enhancement activity.

Clear complaint and redress procedures should be available for cases where students or staff experience discrimination or disadvantage linked to AI systems. Responsibility for coordinating responses and embedding lessons learned should be clearly assigned within existing equality, diversity, and inclusion structures. Students affected by AI-related discrimination should have access to the same support and remediation processes as in other discrimination cases, including academic appeals where AI bias may have influenced assessment outcomes.

Student partnership structures should be embedded in AI governance. This can include student representation on relevant oversight committees, opportunities for student-led evaluation projects, and mechanisms for student bodies to provide feedback on AI experiences and propose equity improvements. These partnership structures ensure ongoing accountability to those most affected by AI adoption and provide early identification of emerging equity issues.

Regular reporting on equity metrics should be made transparent, demonstrating institutional commitment to transparency and continuous improvement. Where inequities are identified, institutions should outline planned actions and timelines for addressing them. Sharing outcomes and learning across the sector through existing national coordination forums supports collective improvement and prevents repetition of mistakes.

2.2.9 Summary of Recommendations

(1) Publish an explicit equity and inclusion commitment in institutional AI policy, aligned with Irish equality law, the Public Sector Equality and Human Rights Duty, and SDG4.

(2) Operationalise equity commitments through procurement, staff and student development, and curriculum design, explicitly recognising intersectional disadvantage.

(3) Provide equitable access to AI tools by securing institutional licences to approved systems so access does not depend on students' private means.

(4) Ensure cross-disciplinary AI infrastructure on campus, including reliable broadband, hardware, and software, with capacity to provision access for individual students where needed.

(5) Address broader digital exclusion through support measures such as extended study spaces, flexible AI literacy training, and mobile-optimised tools.

(6) Support sector-wide collaboration and information sharing to reduce costs, ensure compliance, and promote consistent criteria for approved tools.

(7) Ensure assessment equity by specifying institutionally provided AI tools or requiring transparent declarations where private tools are used.

(8) Adapt assessment rubrics to evaluate critical engagement with AI, not sophistication of outputs, and prioritise assessment strategies that minimise advantage from differential access.

(9) Provide controlled AI access in time-bound assessments to maintain fairness under standardised conditions where AI is permitted.

(10) Apply explicit equity criteria in procurement and approval of AI systems, requiring evidence of data representativeness, governance, and limitations.

(11) Adopt a precautionary approach to systems that lack transparency about training data, risks, or reliability, deferring approval until adequate assurance is available.

(12) Undertake pilot testing with diverse student cohorts prior to large-scale adoption, documenting outcomes in transparent approval reports.

(13) Evaluate AI systems for Irish-language capability and put in place compensatory supports or alternative arrangements where performance is inadequate.

(14) Work collectively across the sector to strengthen Irish-language functionality in AI systems and to support the development of solutions that ensure equitable access for Irish-language education.

(15) Extend language equity measures to minority and migrant languages, ensuring students are not excluded from AI-enabled learning opportunities.

(16) Position AI as a support within inclusive education strategies, ensuring integration into universal design, inclusive pedagogy, and adequately supported student services.

(17) Provide equitable AI-literacy and professional-development opportunities for all staff, ensuring accessibility across disciplines and employment types and recognition within workload and development frameworks.

(18) Deliver AI literacy for all students across programmes, designed for accessibility, multiple formats, and recognition through badges or micro-credentials.

(19) Integrate equity review within existing quality-assurance processes, publish outcomes disaggregated by relevant demographics, establish clear complaint and redress pathways, and embed student representation within AI-governance structures.

2.3 Critical Engagement, Human Oversight, and AI Literacy

Critical engagement, human oversight, and AI literacy form the pedagogical and intellectual foundation for responsible AI integration in higher education. This principle recognises that the transformative potential of generative AI can only be realised through cultivated discernment rather than passive adoption or reflexive rejection. Human oversight is a regulatory requirement, but it is also an epistemological necessity for the preservation of academic judgement in an era where gen AI can mimic scholarly production.

Critical engagement requires institutions move beyond instrumental questions of how to use AI towards fundamental inquiries about when and why such use serves educational purposes. AI literacy, in this context, transcends surface-level tool competence to encompass both conceptual understanding and critical analysis. Students and educators must be able to situate AI in broader debates about knowledge creation, the social implications of algorithmic mediation, and the ethical dimensions of human–machine collaboration. This form of literacy also depends on a foundational level of technical understanding that enables informed and responsible participation in these discussions.

Educators cannot meaningfully critique the epistemic limits of gen AI without grasping, even in simplified form, how architectures such as transformers represent and process language, or how training data and model design introduce bias and constraint. Without such knowledge, critique risks becoming rhetorical rather than substantive. Accordingly, AI literacy in higher education should be understood as layered.

It requires a foundation of technical understanding sufficient to explain how generative models operate in theory. It demands a capacity to apply that knowledge when evaluating AI outputs in teaching and research, alongside a set of critical faculties that enable the contextualisation of algorithmic suggestions within disciplinary standards and societal values.

The imperative for human oversight stems from recognition that educational decisions carry moral weight that cannot be delegated to systems lacking comprehension of their consequences.

2.3.1 Embedding AI literacy as core competency

Programmes within higher education should integrate AI literacy as a core graduate attribute. This requires reconceptualising curricula so that critical understanding of AI is positioned alongside traditional disciplinary knowledge. The European Digital Competence Framework for Citizens (DigComp) emphasises that digital literacy must include not just functional skills but also critical awareness of technology's societal implications.[22] For generative AI, this means a comprehensive grasp of both technical foundations and ethical dimensions.

Institutions should define programme-specific learning outcomes that articulate how AI literacy is expressed within each discipline. These outcomes should include a conceptual understanding of how generative AI processes information and produces outputs, covering, for example, statistical prediction, training data, and model architectures. Students should develop disciplinary awareness of when and how AI is relevant, together with critical-evaluation skills to judge accuracy, bias, appropriateness, and alignment with disciplinary standards.

22 'Digital Competence Framework for Citizens (DigComp).'

They should also cultivate ethical reasoning concerning authorship, accountability, equity, and the preservation of human judgement, as well as practical capabilities in prompt design, verification, and collaborative human–AI workflows.

Generic digital-skills provision alone is insufficient. A medical student must understand how AI diagnostic tools interact with clinical judgement and patient care ethics, an engineering student requires knowledge of how AI-generated designs must be validated against safety standards and professional liability, and a humanities student must engage with how AI text generation relates to authorial voice, interpretative traditions, and cultural production. For this reason, AI literacy should therefore be scaffolded across the stages of a programme, progressing from introductory awareness to advanced critical application, with definitions of 'advanced' determined by teaching staff within their disciplinary context. This progression depends on robust, discipline-specific AI literacy training for staff, ensuring that educators have the knowledge and confidence to define and assess critical engagement in their fields.

At introductory levels, students might engage the technology through guided reflection and low-stakes experimentation, comparing their own reasoning with AI outputs and identifying limitations or errors. At intermediate stages, analysis could extend to comparing responses across multiple systems, experimenting with prompt design, and exploring how training data influences interpretation in their discipline. By later stages, students should demonstrate sophisticated critique and application, whether through research on AI's professional implications, applied use cases, or evaluation of the epistemological consequences of machine-generated outputs. Programme teams should ensure that this progression is coherent and deliberate, with learning outcomes and assessment strategies mapped to promote cumulative development.

Students should also encounter AI through interdisciplinary perspectives that highlight its wider societal significance. Cross-programme seminars, collaborative projects, and integrated modules should bring together perspectives from technology, humanities, social sciences, and professional studies. Joint teaching initiatives and guest lectures from diverse practitioners and researchers should, where feasible, be embedded within formal curricula rather than confined to optional enrichment. Embedding both disciplinary depth and interdisciplinary breadth ensures that graduates emerge not only with functional skills but with the conceptual, critical, and ethical understanding required to navigate an AI-saturated world.

2.3.2 Human oversight in pedagogical processes

The EU AI Act's designation of educational AI as high-risk systems mandates human oversight, recognising that educational decisions shape human potential in ways that require moral accountability. Article 14[23] requires oversight mechanisms that minimise risks to fundamental rights, ensure systems are used as intended, and allow human intervention or discontinuation when necessary.

This means academic staff should retain ultimate responsibility for pedagogical decisions. This principle is most critical in assessment, where grading requires professional judgement that algorithms cannot replicate. Decisions about whether work meets the required standard, or how effectively a student has demonstrated

23 https://artificialintelligenceact.eu/article/14/

understanding, must rest with academic staff who can consider disciplinary context and individual circumstances. Institutions should maintain clear protocols to ensure that all evaluative decisions about student work are made by humans, preserving the essential role of academic expertise in recognising and validating learning.

Comparable oversight is needed in curriculum design. Adaptive learning platforms that promise personalisation may inadvertently narrow intellectual horizons. When algorithms optimise for engagement or completion rates, they may steer students away from difficult concepts or controversial topics that are essential to disciplinary understanding. Academic staff should therefore review algorithmic recommendations against programme learning outcomes to ensure that systems do not limit exposure to the full breadth of disciplinary knowledge.

Effective oversight should be embedded within existing academic-governance structures. Institutions may designate committees or roles responsible for monitoring the use of generative-AI systems, ensuring compliance with oversight principles, and recommending suspension where systems fail to meet required standards. Disciplinary units could identify coordinators with understanding of both the technological and pedagogical dimensions of AI, enabling them to advise colleagues, audit local practice, and raise concerns through established reporting routes.

Oversight structures should have clearly defined responsibilities and the authority to act on identified risks, including requesting documentation, commissioning bias testing, and recommending changes to practice. Locating such authority within recognised governance frameworks ensures that oversight is substantive rather than symbolic, enabling institutions to benefit from AI's analytical capacities while safeguarding the human expertise and ethical judgement that define higher education.

2.3.3 Developing critical AI engagement

Developing critical engagement with AI requires intellectual frameworks that go beyond operational competency. Students must not only learn how to use AI tools effectively but also acquire the ability to interrogate their implications and limitations. This literacy draws on philosophy, sociology, linguistics, and postcolonial critique, interrogating intelligence and consciousness, analysing power and truth, and exposing whose knowledge is privileged. Institutions must ensure students gain a conceptual vocabulary that frames gen AI as a sociotechnical system rather than a neutral tool, recognising how datasets encode historical inequities, how model architectures reflect design choices, how deployment contexts shape interpretation, and how feedback loops can amplify bias. Students should understand gen AI systems as cultural artefacts, marked by the values and blind spots of their creators and training data.

Such engagement requires historical perspective. AI must be situated within the longer history of automation and social change. Students should explore how earlier technologies promised liberation but sometimes introduced new forms of control; how efficiency gains have often accrued unevenly; and how technological determinism can obscure political choice. This perspective supports critical reflection while leaving room for responsible innovation. Disciplinary perspectives enrich this analysis: philosophy addresses agency and moral responsibility, sociology and anthropology reveal AI's embeddedness in social relations, literary and cultural studies analyse its impact on creativity and representation, economics and political science show how AI reshapes labour markets and democracy. Together, these lenses foster comprehensive understanding of gen AI's implications.

Critical engagement also depends on robust evaluative practices. Students must move beyond binary acceptance or rejection of AI outputs and instead apply systematic protocols suited to their fields. Empirical disciplines may emphasise verification of sources and identification of data artefacts, while interpretive disciplines can focus on nuance and rhetorical coherence. All students should be equipped to detect bias, analyse how prompting influences outputs, and understand that bias operates not only technically but structurally, requiring broader societal awareness.

For this vision to be realised, teaching staff need structured opportunities for professional development. Institutions should provide accessible development that builds shared foundations across disciplines, covering technical fundamentals, pedagogical implications, ethical and policy frameworks, and approaches to course integration. Progression pathways should allow staff to develop expertise in areas such as assessment design, AI ethics, technical application, or educational research, recognised through institutional or sectoral accreditation. Peer-learning networks and communities of practice can sustain development across disciplines and institutions, building collective expertise.

Professional development should also address resistance and anxiety. Many educators experience uncertainty about AI or perceive it as incompatible with their professional values. Development programmes should provide space for dialogue, acknowledge concerns about workload and identity, and position AI as a tool that supports, rather than replaces, academic judgement. Low-stakes experimentation and recognition of diverse approaches can build confidence incrementally. Values-based framing should link gen AI integration directly to educational purposes, such as enhancing learning, preserving rigour, and freeing staff to focus on high-value interactions.

Effective adaptation requires time and institutional support. Building AI literacy prior to curriculum or assessment redesign should be recognised within workload and development frameworks rather than treated as an additional burden. Institutions should integrate this activity into existing professional-development systems, ensure equitable access for part-time and sessional staff, and provide access to pedagogical and technical guidance. Incentives for innovation and recognition of exemplary practice can further embed a culture of critical and ethical engagement with AI across the sector.

2.3.4 Student Development Pathways

Developing AI literacy among students requires coherent and progressive curriculum design. Institutions should establish clear pathways that ensure all students achieve threshold competencies while creating opportunities for advanced development among those pursuing AI-intensive disciplines or careers. Foundational learning should be provided early in programmes, introducing conceptual understanding of generative AI, critical evaluation frameworks, ethical considerations, institutional policies, and practical skills for appropriate use.

Progressive integration across programmes should embed AI literacy into disciplinary contexts. This progression should be mapped at programme level to ensure coherence, coverage, and the avoidance of redundancy. Co-curricular opportunities can extend learning for those who wish to specialise, through AI literacy certificates, hackathons, research assistantships, and peer tutoring. These should be accessible through varied formats, times, and entry points to promote inclusion.

Pathways should accommodate the diversity of students' backgrounds and capabilities. Students enter higher education with different levels of confidence in generative AI, from those already programming models to those with limited digital experience. Differentiated entry points can recognise prior learning while guaranteeing critical engagement for all. Importantly, coding ability cannot substitute for critical thinking, and all students should demonstrate the ability to evaluate and contextualise AI outputs.

Accessibility must be built into all provision. Students with learning differences may require additional support with abstract concepts or alternative assessment approaches, while international students may require language support for technical vocabulary and cultural contextualisation of AI examples.

AI literacy also entails ethical development and digital citizenship. Students should be prepared as responsible participants shaping gen AI's societal impact. Academic integrity education should help students to understand why intellectual effort matters and what responsibilities accompany the use of generative AI. Case studies, discussions of ambiguous scenarios, and reflection on personal values can develop nuanced understanding beyond rule compliance. Students should also build data consciousness, recognising how their interactions contribute to training, what rights they have over their data, and how to evaluate privacy practices. They should be encouraged to consider social responsibility, how gen AI adoption may reshape employment in their field, what professional obligations exist for transparency, and how they can contribute to beneficial and ethical AI development. Graduates should emerge not only able to use AI but equipped to shape its direction in ways consistent with societal needs and the values of higher education.

2.3.5 Institutional infrastructure for AI literacy

Effective AI literacy development depends on institutional structures that extend beyond fragmented departmental initiatives. Central coordination helps to ensure consistency, coherence, and access to expertise that individual units may not sustain alone. Institutions should consider designating a coordinating unit or network as a focal point for staff and student development in AI literacy, supported by educational developers and technical specialists. Such provision should offer curriculum resources, coordinate professional development, advise on integration, and evaluate effectiveness. Accessibility should be ensured through a combination of physical and virtual spaces that enable collaboration, workshops, and responsive support.

Resource development should be systematic and sustainable. Discipline-specific exemplars, assessment templates, and interactive tutorials should be developed in partnership with academic departments, regularly updated to reflect technological change, and maintained through version control and periodic review. Quality assurance should be integrated into existing institutional frameworks, including evaluation of learning outcomes, feedback from staff and students, benchmarking against sector practices, and continuous enhancement.

A reliable technological environment is essential for authentic learning and safe experimentation. Institutions should provide licensed access to approved gen AI tools, sandbox environments for exploration, and controlled API access for advanced users. These platforms should operate under clear usage policies, data protection and security protocols, and responsive support.

Partnership and collaboration are central to sustainable provision. Industry links can offer students authentic experience through guest lectures, internships, and project-based collaboration, while institutions retain academic independence and critical distance. Inter-institutional cooperation allows sharing of resources, joint staff development, and collaborative research on effective practice, with sector-level coordination enhancing collective impact. Engagement with the wider community through public lectures, school outreach, and lifelong-learning initiatives extends the benefits of AI literacy beyond higher education and reinforces the sector's civic role in shaping responsible technological futures.

2.3.6 Evaluation of AI Literacy

Assessment of AI literacy should evaluate technical proficiency alongside critical thinking and ethical reasoning. Traditional testing of knowledge recall or procedural skills cannot capture the judgement that genuine literacy demands.

A range of assessment approaches may be appropriate. Portfolios can allow students to demonstrate development over time through collected artifacts, reflective commentary, and evaluation of both their own and others' gen AI use. Case-based assessments present complex scenarios where students must weigh benefits and risks of gen AI deployment or resolve ethical dilemmas. These tasks reflect authentic professional contexts rather than abstract exercises. Collaborative projects add a social dimension, requiring students to work in groups to investigate AI's implications or use gen AI tools responsibly, with individual contributions documented through process notes and peer evaluation.

Evaluation of AI literacy should also occur at programme level. Institutions should consider whether graduates achieve threshold competencies through mechanisms such as capstone projects, cumulative portfolios, or external validation by employers and professional bodies. Results from such evaluations should inform ongoing programme review, highlighting gaps, revising outcomes, and guiding resource allocation. In certain disciplines, employer feedback is particularly important, and should be gathered systematically through focus groups and analysis of placement data to ensure graduates meet professional expectations. Longitudinal studies of student progression can reveal how literacy develops over time, identifying effective practices and areas for enhancement.

Given the rapid evolution of AI technologies, AI literacy provision should operate within continuous-improvement frameworks. Regular review, piloting, and the transparent dissemination of all findings guard against stagnation and ensure that AI literacy education remains responsive to technological and societal change.

2.3.7 Governance and Accountability

Effective implementation of critical engagement, human oversight, and AI literacy requires both visible leadership and distributed responsibility. Ultimate accountability rests with senior leadership, while operational responsibility is embedded across all levels of the institution. Executive sponsorship signals priority and provides external advocacy. The senior sponsor should chair institutional AI governance, reporting to academic council to ensure alignment with the educational mission.

Academic leadership through heads of school or unit ensures disciplinary integration, with responsibility for adapting policy to local contexts, supporting staff development, and monitoring course transformation. These leaders act as bridges between institutional strategy and departmental practice, translating policy into pedagogy and ensuring coherence between frameworks and classroom implementation.

Operational coordination depends on clearly defined roles. Institutions may designate an AI-education lead to oversee literacy development and service coordination, supported by local coordinators who provide discipline-specific advice and escalate emerging issues. These roles should be recognised through workload allocation, administrative support, and inclusion within performance and development frameworks.

Institutions should establish proportionate metrics that capture compliance, effectiveness, and risk. Implementation indicators might track the proportion of programmes with integrated AI literacy, the share of staff completing professional development, and the number of students achieving threshold competencies. Quality indicators should measure student confidence and staff satisfaction. Risk indicators should identify potential problems such as integrity breaches, student complaints, or technical failures. Regular reporting, both internal and public, provides transparency through annual summaries of progress and case studies of innovative practice. Independent review through external experts, benchmarking, and sector-wide participation validates standards and identifies areas for improvement.

Sustainability requires strategic planning and reliable provision of core supports. Institutions should plan for long-term investment in infrastructure, staffing, licensed tool access, and professional development. Efficiency can be achieved through shared services, scalable delivery of foundational content, and peer learning networks, provided that quality and equity are maintained. Enduring impact will depend on combining clear leadership, distributed responsibility, continuous evaluation, and sustainable planning within a coherent institutional framework.

2.3.8 Summary of Recommendations

(1) Recognise AI literacy as a core graduate attribute in all programmes.

(2) Define programme-specific learning outcomes covering technical foundations, disciplinary applications, critical evaluation, ethics, and practical workflow skills.

(3) Map a scaffolded progression from introductory awareness to advanced critical application across stages of study, with assessment aligned to outcomes.

(4) Provide discipline-specific professional development opportunities so educators can teach and assess AI literacy with confidence and credibility.

(5) Embed interdisciplinary perspectives through seminars or modules linking technical, social, ethical, and cultural perspectives.

(6) Codify human oversight so that academic staff retain final authority over assessment, grading, feedback, and curriculum decisions. Institutions should also ensure that students achieve programme-level AI-literacy outcomes even where specific modules restrict AI use.

(7) Implement oversight protocols ensuring that all machine outputs require human review before action.

(8) Monitor adaptive and personalised systems to prevent curricular ‘filter bubbles’, comparing algorithmic recommendations with programme outcomes.

(9) Establish or designate an institutional oversight committee empowered to request documentation, commission bias testing, pause deployment, and discontinue non-compliant tools.

(10) Identify unit-level AI-education coordinators to advise colleagues, monitor local use, and escalate concerns.

(11) Include oversight requirements in procurement, covering human-in-the-loop capability, intervention and override functions, logging, declarations of fitness for purpose, and known limitations.

(12) Develop student development pathways that provide foundational induction for all learners and optional advanced tracks, projects, and credentials for specialisation.

(13) Differentiate entry points and ensure accessibility for varied prior experience, disabilities, and language needs.

(14) Provide central coordination for AI-literacy resources, exemplars, and responsive support for staff and students.

(15) Maintain secure technical infrastructure: licensed tools, sandbox environments, LMS integration, APIs where appropriate, supported by clear policies and support.

(16) Acknowledge staff time for training, assessment redesign, and course transformation within workload and development frameworks, ensuring equitable provision for part-time and sessional staff.

(17) Assess AI literacy authentically using portfolios, case-based tasks, and collaborative projects aligned with disciplinary contexts.

(18) Evaluate outcomes at programme level using cumulative evidence and, where appropriate, input from employers or professional bodies.

(19) Embed continuous-improvement processes with regular review of content, pilot initiatives, evaluation findings, and transparent dissemination of results

(20) Measure and report progress on implementation, quality, and risk through annual internal and public reporting that includes student perspectives.

2.4 Privacy & Data Governance

Privacy and data governance form a central pillar of responsible AI adoption in higher education, recognising the particular vulnerabilities that arise when educational data interacts with gen AI systems. Learning generates sensitive data, and any perceived efficiency gains from gen AI must not come at the expense of student privacy, autonomy, or trust.

Ireland's strong data-protection tradition establishes a high baseline of expectation. HEIs should act as exemplary custodians of the digital traces entrusted to them, embedding privacy-by-design across procurement, governance, and pedagogy. Security and transparency should extend across the full lifecycle of AI adoption, from initial approval of tools to their everyday use by staff and students.

2.4.1 AI sovereignty

Institutions should ensure that their use of artificial intelligence preserves institutional sovereignty: the ability to retain meaningful control over their own data, systems, and decision-making. AI sovereignty goes beyond regulatory compliance. It is the condition that allows HEIs to act in accordance with their missions, unconstrained by the commercial or technical architectures of external vendors.

AI sovereignty requires that institutional data, intellectual property, and records remain under the stewardship of the institution at all times. Vendor contracts should make explicit that data provided to or generated by AI systems belongs to the institution, that it will not be reused for training without explicit agreement, and that it can be retrieved in standard formats on termination of service. Institutions should have clear exit strategies for all critical systems, so that they can discontinue the use of a tool without losing access to, or control over, institutional records or student data.

Procurement processes should consider sovereignty alongside functionality, cost and accessibility. Preference should be given to systems that rely on open standards, that provide full documentation, and that support interoperability and portability of data. Where reliance on a proprietary model is unavoidable, institutions should document the risks of lock-in and ensure that mitigations are in place, such as time-limited contracts and contingency planning.

Sovereignty also entails the ability to audit and explain the behaviour of systems used in the name of the institution. AI systems must be subject to institutional oversight, and institutions should retain the right to suspend or discontinue a system where its outputs cannot be justified to students, staff, regulators, or the public. This principle is inseparable from academic freedom: teaching, research, and administration must not be dictated by opaque gen AI models or commercial interests that override institutional judgement.

The global concentration of generative-AI development within a small number of large providers presents strategic risks for higher education. Reliance on external proprietary infrastructure may limit institutional and sectoral autonomy in areas such as curriculum design, research direction, and graduate capability development. It can also expose institutions to potential changes in access, pricing, or terms of service that conflict with academic priorities or national policy objectives. Managing these dependencies requires proactive assessment of vendor relationships, transparent contractual arrangements, and the capacity to transition to alternative systems where necessary to protect academic independence and continuity of service.

Generative-AI tools inevitably reflect the priorities, assumptions, and biases of their developers, shaped by the datasets, design decisions, and commercial objectives that underpin them. When adopted at scale without appropriate oversight or contextual adaptation, such systems risk narrowing curricular diversity and privileging particular perspectives. Preserving educational integrity and cultural relevance therefore requires that institutions retain the capacity to adapt, contextualise, or where necessary decline the use of AI systems that do not align with their educational missions or the values of their communities.

2.4.2 Data protection

All HEIs must adopt institution-wide policies that guarantee compliance with GDPR and the EU AI Act, or other applicable data protection and AI-related legal frameworks where they apply. Legal requirements establish the minimum threshold for lawful operation, and institutions are responsible for ensuring their practices meet all relevant statutory obligations.

Beyond compliance, institutions should implement best practices that reflect the sensitivity of educational data. Personal data collection and processing in the context of generative AI should be strictly necessary for clearly defined educational purposes, supported by documented justification for each category of data collected. The lawful basis for processing should be explicit, proportionate, and subject to regular review. The principle of data minimisation should guide all decisions regarding information gathering and retention.

Secure storage and handling of educational data require safeguards commensurate with the sensitivity of the information held. Measures such as encryption, role-based access controls, multi-factor authentication, and comprehensive logging and monitoring represent examples of appropriate controls that can help ensure auditability and resilience. Retention periods should align with statutory requirements and be enforced through automated technical processes wherever possible. Regular security testing, vulnerability assessments, and supplier reviews can provide assurance that controls remain effective against evolving threats, with clear procedures for addressing any findings.

Transparency obligations should extend beyond legal notification to meaningful communication. Staff and students should receive plain-language explanations about data collection purposes, legal bases, categories of data, recipients, international transfers, and foreseeable consequences, both at the point of collection and throughout the lifecycle of their data. Explanations of automated processing should avoid legal jargon while remaining accurate and comprehensive, and should be reviewed periodically to ensure continued accuracy.

Data-subject rights must be practically exercisable and supported through clear institutional processes. Institutions should maintain accessible procedures for individuals to request access to, correction of, restriction of, or deletion of their personal data, and for exercising the right to data portability where applicable. Response times should reflect the significance and urgency of the request. Where technical or legal constraints limit full compliance, institutions should offer proportionate alternatives and ensure that the right to object to specific processing purposes is respected without disadvantage to the individual's educational experience.

2.4.3 Impact assessments and special categories

Data Protection Impact Assessments (DPIAs) should be undertaken before the deployment or significant modification of any AI system that processes personal data in educational contexts. Assessments should explicitly consider algorithmic fairness, explainability requirements, data quality, and cumulative privacy risks arising from the interaction of multiple systems or datasets. Reviews of existing DPIAs should capture changes in processing operations, emerging risks, and evolving regulatory guidance. Where residual risks remain high after mitigation, consultation with the relevant supervisory authority should be undertaken.

Special categories of personal data require safeguards reflecting their heightened sensitivity. Health data, disability accommodations, religious or philosophical beliefs, and ethnic origin data that arise in educational contexts demand explicit consent or another appropriate lawful basis with documented necessity and proportionality. Processing should be restricted to authorised personnel with appropriate training, and technical and organisational measures should prevent unauthorised access or disclosure. Regular audits should verify that enhanced protections are consistently implemented.

Cross-border data transfers add further complexity and should be managed with particular care. Where gen AI systems involve international data flows, institutions should ensure that appropriate safeguards are in place. Transfers to jurisdictions without adequate data-protection standards should be subject to specific risk mitigations, and data localisation should be considered, where feasible, for sensitive educational records and special-category data.

2.4.4 Approved tools and vendor governance

AI tools used in teaching, learning, or assessment should be adopted only following institutional approval processes that verify comprehensive compliance. Reviews should document GDPR and EU AI Act adherence, including data flows, processing locations, sub-processor arrangements, and vendor roles and responsibilities. Vendors should demonstrate recognised security standards, provide breach-response procedures that meet statutory requirements, and enable independent verification of their controls.

Contracts should explicitly protect institutional, staff, and student interests. Data ownership remains with the institution, and educational data should not be used to train external models without explicit and informed consent. Retention periods should be clearly defined and enforced through technical measures, and the secure deletion of data upon contract termination should be auditable. Indemnification and limitation-of-liability provisions should allocate risk appropriately and safeguard institutions from vendor non-compliance or security failures.

Technical documentation should include meaningful disclosure of model purpose, inputs, outputs, update cadence, known limitations, and where legally and commercially permissible, the provenance and characteristics of training data. Vendors should outline bias-mitigation approaches, safeguards against common failure modes such as hallucination or prompt injection, and the mechanisms that enable effective human oversight. Version control and change-logging procedures should allow institutions to track updates that may affect output characteristics or privacy risk.

Approved AI tools should be reviewed periodically to ensure continued compliance and fitness for purpose. Re-evaluation should consider incident history, material model changes, and newly identified risks. Institutions should retain the authority to suspend or discontinue tools where risks become unacceptable or contractual terms change materially. Transition planning should prioritise student fairness, preserve assessment integrity, and ensure continuity of provision.

2.4.5 Transparency and intellectual property

Institutions bear responsibility for ensuring that the operation of generative-AI systems used in teaching, learning, and assessment is transparent and intelligible. Students and staff should be able to understand how approved tools process data and how outputs are generated. Technical documentation should be accompanied by accessible explanations of model purpose, capabilities, limitations, and common failure modes. Institutions should communicate any material changes or newly identified limitations in a timely manner and provide clear channels for staff and students to question or challenge AI-influenced outcomes.

Intellectual property and data rights require careful definition. Student work should not be used for external model training without explicit, informed, and freely given consent. Consent should specify purpose, duration, scope, and procedures for withdrawal, and should not be bundled into general service terms. Students retain full intellectual-property rights to their original contributions, while any AI assistance should be acknowledged according to institutional attribution standards without diminishing student authorship or accountability.

Institutional policies should also address copyright and attribution in AI-mediated contexts. Proportionate review mechanisms should be in place to identify potential infringement, supported by clear guidance on citation, permissible AI use, and disciplinary exemplars. The boundary between appropriate AI support and academic misconduct should be explicitly defined and communicated, with examples relevant to each field of study. Policies should be reviewed periodically to remain aligned with technological developments and evolving legal interpretation.

Protecting student autonomy requires that consent mechanisms provide genuine choice rather than implied acceptance. Requests for consent to process student work through AI systems, to incorporate student data into training datasets, or to participate in AI-mediated assessment should be specific and granular, addressing each distinct purpose and category of processing separately. Consent should be sought at appropriate decision points and presented in clear, accessible language.

2.4.6 Summary of Recommendations

(1) Adopt institution-wide AI data policies that ensure compliance with GDPR, the EU AI Act, and other applicable legal frameworks.

(2) Ensure data collection is strictly necessary for defined purposes, with explicit lawful basis and proportional justification.

(3) Apply the principle of data minimisation, collecting and retaining only what is essential and supported by documented rationale.

(4) Maintain robust security controls, such as encryption, access management, multi-factor authentication, and least-privilege principles.

(5) Align retention periods with statutory requirements and, where possible, enforce them automatically through technical measures.

(6) Conduct periodic security testing and supplier audits, implementing remediation within defined timelines.

(7) Provide plain-language transparency notices at the point of data collection and throughout processing, updated regularly to reflect actual practice.

(8) Ensure that data-subject rights, including access, correction, portability, restriction, erasure, and objection, are readily exercisable without undue barriers.

(9) Respect objections to processing and, where feasible, consider non-AI alternatives without disadvantage to the individual.

(10) Complete Data Protection Impact Assessments (DPIAs) before deploying or significantly changing AI systems, and review them regularly.

(11) Consult the relevant supervisory authority when residual risks remain high after mitigation.

(12) Apply heightened safeguards for special category data, including consent, role-based access, and regular audits.

(13) Manage cross-border data transfers lawfully with appropriate risk mitigations, prioritising localisation for sensitive records where practical.

(14) Approve only AI tools that have undergone institutional review covering data-protection and AI Act compliance, data flows, vendor security, and breach procedures.

(15) Include contractual protections that preserve institutional data ownership, prohibit external model training without consent, enforce secure deletion, and allocate liability appropriately.

(16) Require meaningful vendor documentation describing model purpose, inputs, outputs, training data provenance, limitations, and safeguards.

(17) Re-evaluate approved tools periodically, considering material changes and emerging risks, and retain authority to suspend use where necessary.

(18) Provide accessible explanations of AI system operations, outputs, limitations, and failure modes for staff and students.

(19) Protect student intellectual property and autonomy, ensuring that student work is not used for external model training without explicit, granular consent.

(20) Prohibit fully automated high-stakes decisions and ensure that consequential AI outputs are subject to meaningful human review.

2.5 Sustainable Pedagogy

Sustainable pedagogy encompasses both environmental and educational dimensions, recognising that the computational demands of gen AI systems carry ecological impacts that must be balanced against their benefits.

More broadly, sustainable pedagogy requires that gen AI adoption must enhance rather than undermine the long-term vitality of educational practice. It requires that gen AI tools support the development of enduring intellectual capabilities rather than replacing human judgement, creativity, and critical reasoning.

2.5.1 Environmental sustainability and resource consumption

The environmental impact of generative-AI systems should be considered within higher education's broader commitment to sustainability, recognising that institutions serve as both exemplars and educators in promoting responsible practice. Training large AI models requires substantial computational resources, and their ongoing operation involves continuous energy use that increases with each query and response. HEIs should evaluate the environmental costs of AI adoption alongside pedagogical benefits, seeking efficient implementations that minimise unnecessary computational demand.

Procurement decisions should include assessment of environmental impact across the full lifecycle of AI systems, from development and deployment to decommissioning. Vendors should be expected to provide transparent reporting on energy consumption, carbon emissions, and the use of renewable energy sources in their data centres. Preference should be given to providers that demonstrate credible commitments to carbon reduction and employ efficient model architectures capable of achieving comparable educational outcomes with lower resource intensity.

Institutional implementation should prioritise shared infrastructure over duplicated systems to reduce overall energy consumption and maximise utilisation rates. Where feasible, local or edge processing may be preferred over cloud-based solutions that involve repeated data transmission. Techniques such as caching frequently accessed content and batching requests can reduce computational overhead without diminishing educational value. Institutions should periodically review the environmental performance of deployed tools to inform decisions about their continued use or replacement.

Education for sustainability should also include understanding the environmental implications of digital technologies. Students should be encouraged to develop responsible AI-use habits that balance learning benefits with ecological impact. Educators are encouraged to consider whether AI-assisted tasks justify the environmental costs involved and to employ alternative approaches where traditional methods achieve comparable outcomes. Embedding digital sustainability within curricula can help graduates navigate technological futures with environmental awareness and responsibility.

2.5.2 Preserving foundational intellectual capabilities

The convenience of AI assistance should not diminish the development of core academic competencies that remain essential for intellectual growth and professional practice. Writing, mathematical reasoning, critical analysis, and creative problem-solving are foundational capabilities that must be cultivated through direct engagement rather than delegated to algorithmic systems. Institutions should support staff in defining where generative-AI tools can appropriately augment human capability and where they risk replacing the productive effort that underpins intellectual development.

Curriculum design should ensure that students develop robust abilities before engaging extensively with generative AI. Sequencing AI integration across programmes requires careful consideration, with earlier stages prioritising the cultivation of human capabilities and later stages exploring more advanced forms of human–AI collaboration. Assessment strategies should confirm that students possess underlying competencies necessary to exercise critical judgement independent of technological assistance.

Faculty development programmes should equip educators to distinguish between productive AI use that extends learning and problematic dependence that limits intellectual growth. Teaching staff should be supported in redesigning courses to leverage AI responsibly while preserving essential learning challenges.

The drive for efficiency should not lead to the elimination of academic difficulty where that difficulty serves educational purpose. Sustained engagement with complexity – through challenge, reflection, and creative problem-solving – remains indispensable to fostering resilience, insight, and independent thought.

2.5.3 Long-term educational ecosystem health

The sustainability of educational practice depends on maintaining diversity in pedagogical approaches and learning pathways. Over-reliance on particular AI systems risks creating monocultures that are vulnerable to technological disruption or systematic biases embedded in dominant platforms. HEIs must consciously preserve methodological pluralism, ensuring that AI augments rather than replaces the rich variety of educational practices that serve diverse learners.

Staff autonomy in pedagogical decision-making should be protected while maintaining coherent institutional approaches to AI integration. Academic freedom includes the right to critically adopt or decline AI tools based on disciplinary requirements and pedagogical philosophy, provided that core learning outcomes are met. Departments and individual educators should have flexibility to determine appropriate AI use within their contexts, while aligning with institutional principles of equity, accessibility, and academic integrity.

The economic sustainability of AI adoption also requires long-term consideration. Subscription-based models that generate ongoing financial obligations should be evaluated in relation to institutional budgets and alternative investments in human teaching capacity. Risks of vendor lock-in associated with proprietary systems can be mitigated through open standards, data portability, and clear exit strategies that protect institutional autonomy. Cost–benefit analyses should account not only for licensing fees but also for indirect costs, including staff training, technical support, and infrastructure requirements necessary for effective and equitable implementation.

2.5.4 Institutional capacity and resilience

Sustainable pedagogy depends on institutional structures that can adapt continuously as AI technologies evolve. Governance mechanisms should balance agility in responding to technological change with stability in upholding educational mission and values. Decision-making processes should incorporate diverse perspectives – including educators, students, technical staff, and educational developers – while maintaining coherence with institutional strategy.

Resilience planning should anticipate potential disruptions such as vendor failure, system breakdowns, or regulatory changes affecting AI availability. Business-continuity arrangements should enable learning and assessment to continue without reliance on specific AI tools, maintaining alternative pathways that preserve learning outcomes. Regular review and stress testing of AI dependencies can help identify single points of failure and guide mitigation through redundancy or alternative provision.

Investment strategies should balance innovation with sustainability, avoiding overcommitment to emerging or unproven technologies while ensuring continued development of institutional digital capability. Pilot initiatives should test new approaches at a manageable scale before full implementation, allowing lessons to be drawn from experimentation without exposing entire cohorts to undue risk. Evidence-based evaluation of pilot outcomes should inform broader adoption decisions, resisting vendor pressure or competitive anxiety that could lead to premature or unsustainable implementation.

2.5.5 Monitoring and continuous improvement

Sustainable pedagogy requires systematic monitoring of how AI integration affects educational quality, student development, and institutional resilience. Longitudinal evaluation of student cohorts should assess whether anticipated benefits of AI-augmented learning are realised without unintended consequences. Learning analytics should consider not only immediate performance but also long-term retention, transfer, and application of knowledge developed with AI assistance.

Regular reviews should evaluate whether the use of generative AI supports or challenges institutional educational missions, with implementation adjusted in response to evidence rather than assumption. Student feedback mechanisms should capture both short-term satisfaction and reflective evaluation of learning effectiveness. Collaborative research and partnerships across institutions can strengthen sector-wide understanding of effective practice and identify transferable models.

Continuous-improvement processes should respond to emerging evidence while maintaining strategic coherence. Annual reviews of AI integration should consider environmental impact, educational effectiveness, and economic sustainability, informing policy refinement and resource planning. Institutional learning from both successes and challenges should be recorded and shared through sectoral coordination mechanisms, contributing to collective knowledge development and continuous enhancement across higher education.

2.5.6 Summary of Recommendations

(1) Integrate sustainability considerations into AI policy, addressing both environmental and educational dimensions.
(2) Conduct environmental impact assessments within AI procurement and adoption decisions.
(3) Request that vendors disclose energy use and carbon footprint, including commitments to carbon reduction and renewable energy sourcing.
(4) Prefer efficient and low-energy AI models where they deliver equivalent educational outcomes.
(5) Prioritise shared institutional infrastructure over duplicated systems to minimise overall energy consumption.
(6) Optimise technical deployment through caching, batching, and local processing where feasible.
(7) Review institutional AI energy use regularly and act on findings to reduce ecological impact.
(8) Embed digital sustainability education into curricula, so that students understand environmental implications of AI use.
(9) Sequence AI integration to ensure that students first develop core competencies in writing, reasoning, and creativity before applying AI tools.
(10) Design assessments to isolate human competencies, ensuring students can demonstrate mastery independent of AI support.
(11) Provide faculty development on distinguishing between productive augmentation and over-reliance on AI tools.
(12) Maintain diversity of pedagogical approaches, avoiding over-reliance on any single platform or system.
(13) Uphold academic freedom in faculty decisions on AI use, consistent with institutional commitments to equity and integrity.

(14) Evaluate long-term financial implications of AI adoption, including licensing, training, and infrastructure.

(15) Mitigate vendor lock-in risks, favouring open standards, data portability, and clear exit strategies in procurement.

(16) Incorporate resilience planning into AI strategy to ensure continuity of learning if specific tools fail or become unavailable.

(17) Periodically review dependencies to identify and mitigate single points of failure in AI provision.

(18) Adopt pilot-first approaches, testing innovations at small scale and using evidence-based evaluation before wider implementation.

(19) Monitor educational outcomes longitudinally to assess retention, transfer, and long-term student development in AI-enabled contexts.

(20) Review AI adoption regularly for environmental, educational, and financial sustainability, adjusting practice based on evidence and sharing lessons across the sector.

3

References

Digital Competence Framework for Citizens (DigComp). https://joint-research-centre.ec.europa.eu/projects-and-activities/education-and-training/digital-transformation-education/digital-competence-framework-citizens-digcomp_en.

Elhussein, Gene, Elselot Hasselaar, Ostap Lutsyshyn, Tanya Milberg, and Saadia Zahidi. 'Shaping the Future of Learning: The Role of AI in Education 4.0.' World Economic Forum, 2024. https://www.weforum.org/publications/shaping-the-future-of-learning-the-role-of-ai-in-education-4-0/.

Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (Review Draft). OECD, 2025. https://ailiteracyframework.org.

Ethics Guidelines for Trustworthy AI. European Commission: Directorate-General for Communications Networks, Content and Technology and High-level Expert Group on Artificial Intelligence, 2019. https://doi.org/10.2759/346720.

Kestin, Greg, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. 'AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting.' *Scientific Reports* 15 (2025). https://doi.org/10.1038/s41598-025-97652-6.

Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. 'GPT Detectors Are Biased against Non-Native English Writers.' *Patterns* 4, no. 7 (2023). https://doi.org/10.1016/j.patter.2023.100779.

Miao, Fengchun, and Wayne Holmes. 'Guidance for Generative AI in Education and Research.' UNESCO, 2023. https://doi.org/10.54675/EWZM9535.

Miao, Fengchun, and Cukurova Mutlu. 'AI Competency Framework for Teachers.' UNESCO, 2024. https://doi.org/10.54675/ZJTE2084.

O'Sullivan, James, and Colin Lowry. 'Ten Considerations for Generative Artificial Intelligence Adoption in Irish Higher Education.' Higher Education Authority, 2025. https://hub.teachingandlearning.ie/genai/ten-considerations-for-generative-artificial-intelligence-adoption-in-irish-higher-education/.

Otterbacher, Jahna. 'Why Technical Solutions for Detecting AI-Generated Content in Research and Education Are Insufficient.' *Patterns* 4, no. 7 (2023). https://doi.org/10.1016/j.patter.2023.100796.

Perkins, Mike, Leon Furze, Jasper Roe, and Jason MacVaugh. 'The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment.' *Journal of University Teaching and Learning Practice* 21, no. 06 (2024). https://doi.org/10.53761/q3azde36.

Public Sector Equality and Human Rights Duty. Department of Public Expenditure, Infrastructure, Public Service Reform and Digitalisation, 2021. https://gov.ie/en/department-of-public-expenditure-infrastructure-public-service-reform-and-digitalisation/organisation-information/public-sector-equality-and-human-rights-duty/.

Varsik, Samo, and Lydia Vosberg. 'The Potential Impact of Artificial Intelligence on Equity and Inclusion in Education.' OECD, 2024. https://doi.org/10.1787/15df715b-en.

Voultsiou, Evdokia, and Lefteris Moussiades. 'A Systematic Review of AI, VR, and LLM Applications in Special Education: Opportunities, Challenges, and Future Directions.' *Education and Information Technologies* 30, no. 13 (2025): 19141–81. https://doi.org/10.1007/s10639-025-13550-4.

Warschauer, Mark, Waverly Tseng, Soobin Yim, et al. 'The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language.' *Journal of Second Language Writing* 62 (2023). https://doi.org/10.1016/j.jslw.2023.101071.

HEA | An tÚdarás um Ard-Oideachas
The Higher Education Authority

Higher Education Authority
3 Shelbourne Buildings
Crampton Avenue
Shelbourne Road
Ballsbridge
D04 C2Y6

Contact
Phone: +353 1 231 7100
Lo-Call Number: 1890 200 637
Fax: +353 1 231 7172
Email: info@hea.ie
www.hea.ie
@hea_irl