

Perspective

From Permission to Pedagogy: The Structured AI-Guided Education Assessment Policy (SAGE-AP) for Generative AI in Higher Education

Mahmoud Elkhodr *  and Ergun Gide

School of Engineering and Technology, CQUniversity, Sydney, NSW 2000, Australia; e.gide1@cqu.edu.au

* Correspondence: m.elkhodr@cqu.edu.au

Abstract

Higher education policy on generative artificial intelligence has developed rapidly, yet much of this development remains stronger on governance, permission, disclosure, and assurance than on pedagogy. Universities increasingly move beyond blanket prohibition by distinguishing between restricted and permitted contexts, requiring acknowledgement of tool use, and introducing verification mechanisms to protect authorship and understanding. However, publicly visible institutional approaches appear less developed in providing structured, student-facing workflows that guide responsible AI engagement during assessment completion. This article, informed by a bounded qualitative document analysis, uses the term *pedagogical middle layer* to describe the process guidance needed between institutional permission settings and academic-integrity or misconduct procedures. Drawing on recent literature and a purposive scan of selected publicly available university policy and guidance documents, the paper argues that current public-facing models are often effective at defining boundaries but less explicit in guiding disciplined, transparent, and defensible forms of human–AI collaboration. In response, the paper presents the Structured AI-Guided Education Assessment Policy (SAGE-AP) as a theoretically grounded policy proposal for AI-assisted assessment, rather than as an empirically validated policy intervention. SAGE-AP frames assessment as a staged process in which students begin from their own understanding, engage with AI critically, document evaluative decisions, refine outputs responsibly, and defend the reasoning represented in the final submission. The paper contributes to institutional policy development by clarifying how permission settings may be complemented by pedagogical process guidance in the generative AI era.

Keywords: generative AI; higher education; assessment policy; conceptual perspective; conceptual review; document analysis; academic integrity; Structured AI-Guided Education Assessment Policy; SAGE-AP; AI-assisted learning



Academic Editor: Michael DeBraga

Received: 27 May 2026

Revised: 14 June 2026

Accepted: 17 June 2026

Published: 22 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid diffusion of generative artificial intelligence (GenAI) has created substantial pressure on higher education institutions to clarify acceptable student use, protect academic standards, and redesign assessment for a changing technological environment (Fütterer et al., 2023; Miao & Holmes, 2023). Initial institutional responses frequently centred on prohibition, caution, and concern regarding academic misconduct, while more recent policy development has become more differentiated through conditional

permission, acknowledgement requirements, assessment redesign, and assurance of learning mechanisms (Luo, 2024; Moorhouse et al., 2023; Xia et al., 2024; TEQSA, 2023, 2025; Bittle & El-Gayar, 2025). Universities increasingly acknowledge that GenAI cannot be addressed effectively through blanket restriction alone and that assessment policy must distinguish between contexts in which AI use may be inappropriate and those in which it may be both legitimate and educationally valuable (An et al., 2025; Francis et al., 2025; Parker et al., 2025; Perkins et al., 2024).

Despite this progress, an important limitation persists in much of the published policy literature and in many publicly visible institutional documents. Policy is often stronger on permission structures, disclosure obligations, and endpoint verification than on student-facing pedagogical processes (An et al., 2025; McDonald et al., 2025; Parker et al., 2025; Perkins & Roe, 2024). In other words, institutions are becoming clearer about what students may or may not do, but public-facing policy documents are often less explicit about how students should work with AI in disciplined and educationally defensible ways during the completion of an assessment task. This leaves a policy-design space between governance and learning design. The argument is not that pedagogical guidance is absent everywhere; richer guidance may already exist in internal learning management system pages, assessment briefs, staff-facing materials, or local teaching practice that was not accessible to this analysis.

This paper argues that higher education requires a more pedagogically explicit policy layer for GenAI-supported assessment. The term *pedagogical middle layer* is used here to denote the student-facing process guidance that connects institutional permission rules with academic integrity and assurance mechanisms. The purpose of this work is to position the Structured AI-Guided Education Assessment Policy (SAGE-AP) as one theoretically grounded response to that need, not as a uniquely effective or empirically validated solution. In this article, SAGE-AP denotes the policy instrument introduced here, while SAGE denotes the broader Structured AI-Guided Education framework from which it is derived. The argument developed here is that AI policy should not be understood only as a boundary-setting instrument. It should also function as an educational model that guides students in transparent, critical, and defensible engagement with AI across the stages of academic work (Bearman et al., 2024; Chan, 2023; Perkins et al., 2024). On that basis, the paper examines the emerging university policy landscape, identifies recurring strengths and limitations, and advances SAGE-AP as a proposed pedagogically structured response.

Five interrelated concerns frame the argument: student perceptions, critical thinking, epistemic trust, learning processes, and ethical AI use. Students report benefits from GenAI for ideation, writing support, personalisation, and research assistance but also express concerns about accuracy, privacy, dependency, authorship, and longer-term learning consequences (Chan & Hu, 2023; Moorhouse et al., 2023; Xia et al., 2024). These concerns converge around the need for students to evaluate AI-generated content, verify claims against disciplinary anchors, calibrate trust, refine work through visible judgement, and understand why some forms of AI assistance are educationally defensible while others are not (Bearman et al., 2024; Choung et al., 2023; Darvishi et al., 2024; Ennis, 2011; Facione, 1990; Ng et al., 2021). A pedagogical policy layer, therefore, matters because responsible AI use is not only a question of permission or acknowledgement; it is also a question of how students are guided to work, decide, verify, reflect, and remain accountable during assessment completion.

The article is guided by three questions: (1) How does the published literature characterise the current direction of GenAI assessment policy in higher education? (2) What forms of permission, disclosure, documentation, verification, and assurance are visible in selected public institutional documents? (3) How can the broader Structured

AI-Guided Education (SAGE) framework be translated into a student-facing policy proposal that addresses the process of responsible AI-supported assessment? In addressing these questions, the paper makes three contributions. First, it distinguishes between the governance layer of GenAI policy, which determines whether and under what conditions AI may be used, and the pedagogical layer, which guides how students should work with AI once use is permitted. Second, it uses a purposive scan of selected publicly available institutional documents to illustrate how current approaches commonly emphasise permission, disclosure, documentation, and assurance. Third, it presents SAGE-AP as a theoretically grounded, student-facing policy proposal for structuring responsible AI-supported assessment through verification, evaluative judgement, visible refinement, reflection, and defended authorship.

The remainder of the paper is organised as follows. Section 2 describes the approach to the policy document analysis and reviews emerging university policy patterns for GenAI. Section 3 explains why the current policy requires a pedagogical middle layer between permission and misconduct. Section 4 traces the movement from the SAGE framework to SAGE-AP and explains its design logic, clause architecture, pedagogical character, and companion materials. Section 5 positions SAGE-AP relative to selected institutional approaches. Section 6 discusses implications for higher education policy and practice. Section 7 discusses limitations and directions for future research and concludes the paper.

2. The Emerging University Policy Landscape for Generative AI

This section establishes the analytical basis for examining the emerging GenAI policy landscape in higher education. It first explains the article's conceptual perspective, design, and bounded document analysis before identifying assessment-relevant policy patterns visible in selected public institutional documents.

2.1. Conceptual Perspective Design and Bounded Document Analysis

This paper is presented as a conceptual article, compatible with the broader family of conceptual review and theory-building papers and informed by bounded qualitative document analysis. It is not a systematic review, scoping review, or prevalence study. This methodological choice follows from the purpose of the article: to develop and position a policy design argument, rather than to aggregate empirical findings statistically or provide an exhaustive map of global university practice. Conceptual articles require an explicit design logic and a clear account of how literature and artefacts are used to support conceptual argumentation (Jaakkola, 2020). Review typologies also distinguish between systematic evidence synthesis and more flexible conceptual, critical, and narrative approaches that are appropriate when the purpose is interpretation, model development, or theory-informed argument rather than exhaustive retrieval (Grant & Booth, 2009; Snyder, 2019; Torraco, 2005).

The underlying orientation is interpretive. Public policy and guidance texts are treated as institutional artefacts that express assumptions, priorities, and policy logics, rather than as neutral indicators from which sector-wide prevalence can be inferred. This approach is consistent with qualitative document analysis, in which documents are reviewed systematically for meaning, emphasis, context, and underlying assumptions (Bowen, 2009). The purpose of the document analysis is therefore not to produce prevalence estimates, rank institutional policies or guidance documents, or provide an exhaustive account of global university practice. Instead, it uses a bounded and purposive document review to identify visible patterns in selected publicly available GenAI assessment policies and guidance documents. This approach is appropriate for the paper's conceptual aim,

which is to position SAGE-AP in relation to emerging policy logics rather than to map the entire policy field.

The document set was selected in April-May 2026 from publicly accessible university policy and guidance pages concerned with student use of generative AI in assessment. Documents were identified through targeted web searches of institutional websites and public search results using combinations of terms such as “generative AI assessment policy”, “AI use in assessment”, “academic integrity generative AI”, and the relevant institution names. Searches were not database-driven because the objects of analysis were public institutional policy and guidance artefacts rather than indexed scholarly articles. Seven institutional examples were included: the University of Sydney, Monash University, the University of Technology Sydney, University College London, City, St George’s, National University of Singapore, and Tsinghua University. These examples were selected purposively to provide cross-jurisdictional coverage across Australia, the United Kingdom, Singapore, and China and to capture variation in policy and guidance emphasis, including assessment architecture, acknowledgement and disclosure, process documentation, task-level guidance, critical checking, authorship verification, privacy, and defended understanding. The selection was guided by a maximum-variation logic intended to capture different policy emphases and jurisdictional traditions rather than to estimate prevalence. Australian examples were more heavily represented because the jurisdiction has been at the forefront of regulator-led assessment reform in response to generative AI, with TEQSA’s (2023, 2025) sector resources establishing a comparatively mature policy environment. United Kingdom examples were included to capture both undergraduate categorisation approaches and the doctoral viva tradition. Singapore and China examples were included to ensure that the scan was not confined to Anglophone Western jurisdictions. The corpus, therefore, reflects deliberate sampling for variation across policy logic rather than coverage of the global sector.

Documents were included where they were publicly available, institutionally authored or institutionally hosted, and directly relevant to GenAI use in student assessment, academic integrity, or higher education learning practice. Documents were excluded where they were limited to general institutional AI strategy, staff-only operational advice, media commentary, technical procurement guidance, or broader digital transformation statements without substantive assessment relevance. The analysis did not include password-protected learning management system materials, internal academic integrity case procedures, or unit-level assessment briefs that were not publicly available.

The selected documents were reviewed using a structured extraction matrix focused on seven dimensions: (1) whether the document specified conditional permission or prohibition; (2) whether it required acknowledgement or disclosure; (3) whether it requested process documentation such as prompts, drafts, or iterations; (4) whether it addressed independent verification of AI outputs; (5) whether it included privacy, confidentiality, or intellectual property requirements; (6) whether it provided student-facing guidance on how to work with AI during task completion; and (7) whether it included an assurance mechanism such as oral explanation, viva-style scrutiny, live demonstration, or other evidence of defended understanding. The analysis was interpretive and comparative. Its purpose was to identify recurring policy or guidance emphases and visible limitations against which the pedagogical contribution of SAGE-AP could be positioned.

The coding procedure operated in two stages. In the first stage, the first author conducted a close reading of each document and applied the seven-dimensional matrix to extract evidence relevant to each analytical category. Extracted material was recorded against each dimension for each institution, producing the populated matrix presented in Appendix A. In the second stage, the second author independently reviewed the extracted material against the original source documents and against the seven dimensions,

noting any disagreement, omission, or ambiguity. Disagreements were reconciled through discussion until both authors were satisfied that the matrix entries accurately reflected the source content. The procedure did not employ a formal inter-rater agreement statistic such as Cohen's kappa, which is consistent with the focused and bounded nature of the scan and with the qualitative and interpretive purpose of the analysis. A summary of the selection scope, inclusion and exclusion criteria, and rationale is provided in Table A1, in Appendix A. The full list of documents reviewed, with titles, sources, and access dates, is provided in Appendix A, Table A2.

Because the document set was purposive and limited, the analysis should not be read as a comprehensive survey of institutional policy or guidance. The corpus is bound to seven publicly available English-language institutional artefacts accessed in May 2026, and no formal inter-rater agreement statistic was calculated. Broader claims are therefore grounded primarily in the published literature, while the selected documents are used illustratively. These methodological boundaries, including the possibility that richer guidance exists in internal learning management systems, staff-only materials, unit-level assessment briefs, or local teaching practice, are returned to in Section 7.

2.2. Emerging Policy Patterns

Recent scholarship indicates that higher education policy on generative artificial intelligence remains in rapid transition. Large-scale policy mapping suggests that institutional responses have become more visible but also uneven, decentralised, and dependent on local instructor judgement. Parker et al.'s comparative analysis of artificial intelligence policies across 343 universities in Australia, Canada, China, the United Kingdom, and the United States found that 52% of institutions left decisions about AI usage to instructors, 46% deferred to instructor approval for AI use in courses, and 39% did not present a stated rationale for their AI policy (Parker et al., 2025). An et al.'s analysis of GenAI guidelines at the top 50 United States universities similarly found that faculty-facing guidance was far more common than student-facing guidance, with academic integrity, checking with instructors, and tool limitations dominant where student guidance existed (An et al., 2025).

Assessment-focused and policy-focused reviews point in the same direction. Moorhouse et al. found that fewer than half of the world's top-ranking institutions had publicly available GenAI assessment guidelines at the time of review, and that available guidance focused mainly on academic integrity, assessment design advice, and communication with students (Moorhouse et al., 2023). Luo's critical review showed that institutional policy discourse frequently frames GenAI through originality, plagiarism, and misconduct (Luo, 2024). Chan's AI Ecological Education Policy Framework, McDonald et al.'s institutional policy analysis, Perkins and Roe's corpus study of academic integrity policies, and TEQSA's assessment reform resources all reinforce the need to treat GenAI policy as a connection between governance, integrity, assessment design, and student-facing learning practice rather than as a detection or disclosure problem alone (Chan, 2023; McDonald et al., 2025; Perkins & Roe, 2024; TEQSA, 2023, 2025).

The selected institutional documents are used here to illustrate, rather than estimate the prevalence of, these policy logics. The Australian examples show movement towards conditional permission, assessment architecture, acknowledgement, and process evidence. The University of Sydney's two-lane assessment model distinguishes secure and open assessment contexts; Monash University's guidance is comparatively developed in requiring summaries of tool choice, use purpose, iterations, and adaptation of outputs; and UTS provides task-level guidance that narrows ambiguity for specific assessment designs (Monash University, 2026; University of Sydney, 2025; University of Technology

[Sydney, 2025](#)). These documents are valuable, but their dominant focus remains permission architecture, disclosure, documentation, and local rule-setting rather than a complete staged student workflow.

The United Kingdom and Asian examples illustrate adjacent emphases. UCL categorises permitted uses of GenAI and requires acknowledgement where relevant, while City St George's preserves viva-style scrutiny in doctoral assessment contexts where authorship and understanding must remain demonstrable ([City St George's, University of London, 2026](#); [University College London, 2026](#)). The National University of Singapore and Tsinghua University provide comparatively mature guidance on acknowledgement, critical checking, oversight, course alignment, and thesis supervision ([National University of Singapore, 2026](#); [Tsinghua University, 2025](#)). These examples show that public-facing policy has moved beyond prohibition and generic warning. However, they also illustrate that verification, disclosure, oversight, and defended understanding often appear as separate mechanisms rather than as a unified student-facing process.

In sum, the published literature and selected public documents suggest a more specific gap than a simple absence of policy. Current policy and assessment frameworks are increasingly useful for determining whether AI may be used, under what conditions, how AI use should be acknowledged, and how learning may be assured. Fewer public-facing artefacts provide a concise workflow for how students should engage with AI throughout the assessment process once use is permitted. In concrete terms, most public-facing documents tell students whether they may use AI, how to acknowledge it, and what follows if they misuse it, while fewer provide a structured sequence that begins from the student's own thinking, proceeds through verified engagement, and ends with a defensible submission. SAGE-AP is positioned in relation to this less visible public-facing policy layer as one theoretically grounded policy proposal, not as a unique or empirically validated solution.

3. Why Policy Needs a Pedagogical Middle Layer

A recurring risk in current GenAI policy is that public-facing guidance can oscillate between two poles. At one end lies the question of permission: whether students are allowed to use AI in a given task and under what conditions. At the other lies the question of misconduct: whether AI use becomes excessive, deceptive, or incompatible with authorship expectations. Between these poles lies a pedagogical middle layer: the practical, student-facing process guidance that helps learners work with AI responsibly before a submission reaches the point of assurance or integrity judgement. This layer may exist in local or internal teaching materials not captured in public policy scans, but it remains less visible in many public-facing documents and less developed as a reusable policy form in the published literature.

As established in Section 2, this characterisation should not be read as suggesting that the assessment literature lacks middle-ground responses to GenAI. Frameworks such as the Artificial Intelligence Assessment Scale already help educators specify permitted levels of AI use and align AI conditions with learning outcomes, while assessment reform scholarship and regulatory-sector resources increasingly emphasise evaluative judgement, learning assurance, and responsible AI use rather than detection alone ([Bearman et al., 2024](#); [Perkins et al., 2024](#); [TEQSA, 2023, 2025](#)). The gap identified here is therefore narrower: existing frameworks are increasingly useful for setting assessment conditions and assurance expectations, but they do not always provide a concise student-facing workflow for how learners should proceed once AI use is permitted. This is the specific pedagogical layer to which SAGE-AP contributes.

Figure 1 summarises this relationship.

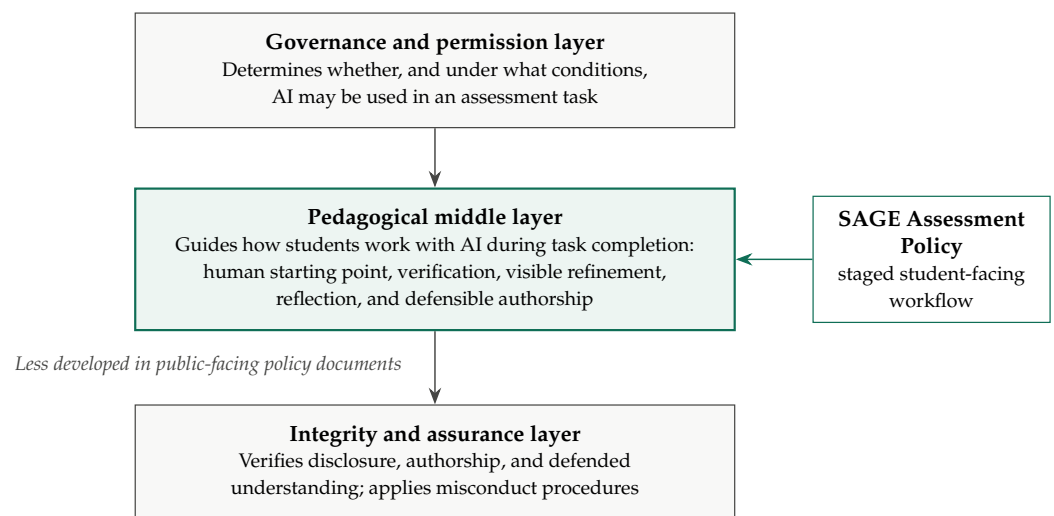


Figure 1. The pedagogical middle layer between permission and misconduct in GenAI assessment policy.

This middle layer matters because responsible AI-supported assessment is not achieved simply by allowing AI use and asking students to disclose it (Kofinas et al., 2025). Students also need guidance on how to begin from their own understanding, how to evaluate AI-generated outputs, how to distinguish useful assistance from misleading or low-quality content, and how to document the reasoning through which they refine and adopt material into a final submission. Without such guidance, policy risks becoming either permissive but educationally shallow or restrictive without providing an alternative.

A pedagogical policy model, therefore, needs to do more than define boundaries. It should translate expectations into an observable process of academic work. This includes an explicit starting point in the student's own thinking, a mechanism for checking AI outputs against disciplinary anchors, a way of recording evaluative judgement, an expectation of reflective refinement, and a means of defending the quality and reasoning represented in the final submission. In this sense, the challenge is not only to manage AI. It is to design a policy that supports the development of judgement, responsibility, and defensible academic agency in AI-rich environments.

4. From the SAGE Framework to SAGE-AP

SAGE has already been introduced in prior publications and earlier sections of this paper, so the aim here is not to reintroduce it as a new model. Rather, this section briefly clarifies the part of the broader SAGE research arc that matters for understanding the assessment policy developed from it. At the framework level, SAGE is a pedagogical process for AI-integrated assessment rather than a permission framework. It structures student work through six linked stages: Generate, Evaluate, Refine, Audit, Reflect, and Defend. In this paper, the term *audit* is retained to emphasise the evaluative function of the fourth stage within the cycle. The essential point is that SAGE addresses how students work with AI inside an already permitted assessment context, whereas institutional policy frameworks more commonly determine whether and how much AI use is allowed in the first place (Elkhodr & Gide, 2026b).

This distinction also clarifies the contribution of the present article relative to earlier SAGE publications. Prior work introduced and examined SAGE primarily as a pedagogical framework for structuring AI-supported learning and assessment. The present paper does

not repeat that contribution and does not claim to provide a new empirical validation of SAGE. Its more specific contribution is to translate that framework into a policy instrument and to locate it within the broader institutional problem of connecting permission settings with student-facing process guidance.

The policy discussed in this paper emerged from that broader framework rather than preceding it. As reported in the prior SAGE research arc (Elkhodr & Gide, 2025), the framework was developed through a sustained program of pedagogical inquiry across multiple Australian campuses and disciplines, including cybersecurity, systems analysis and design, data analytics, and first-year ICT. The full framework documentation is also available as an open repository deposit (Elkhodr & Gide, 2026c). That wider program repeatedly surfaced three pedagogical problems that are directly relevant to policy formulation: students often need explicit structure to verify AI claims against authoritative sources; they need support to move beyond superficial editing toward visible intellectual revision; and open AI use creates an assurance gap between polished submission and demonstrated understanding. These observations are referenced here not as empirical validation of the present policy, which is a conceptual and design contribution rather than a measured intervention, but as the pedagogical context that explains why the policy emphasises verification, visible revision, and a controlled Defend step.

Against that background, SAGE-AP can be understood as a translation of framework logic into a concise, student-facing policy instrument. It is constructed as a principles-based document rather than a permission taxonomy. This distinction is deliberate and responds directly to the gap identified in Sections 2 and 3. Existing institutional policies, whether structured as two-lane models, five-category scales, or six-level frameworks, are predominantly concerned with the question of how much AI may be used in a given assessment (Perkins et al., 2024). They operate at the governance layer. SAGE-AP is intended to operate at the pedagogical layer, specifying how students should engage with generative AI once the permission boundary has been set elsewhere. The two layers are complementary rather than competing, and the policy states this relationship explicitly in its opening scope clause.

Three design commitments follow from this logic. The first is that the policy must be adoptable by educators who do not wish to use the SAGE label. If the principles are to be useful beyond a single branded framework, they must be expressible independently of the branding. The policy, therefore, presents each principle as a standalone statement, with SAGE cycle cross-references provided only where relevant, and is accompanied by a framework-neutral variant that preserves the core expectations without naming SAGE; this variant is presented as a fourteen-principle table in Appendix B and is provided, with the full companion materials, in Supplementary File S1.

The second commitment is that the policy must be directly student-facing rather than educator-facing. Most university AI policies in the landscape reviewed in Section 2 are written as governance documents directed at course convenors or academic managers. Students encounter them, if at all, through abbreviated notices in assessment briefs. SAGE-AP inverts this orientation. It is written in the second person, uses plain language where possible, and is designed to be printed and attached directly to an assessment brief or published in a learning management system.

The third commitment is that each clause must carry both a rule and a justification. This responds to the design assumption that rules are more educationally useful when their purpose is explicit and when students can see how the requirement relates to learning, authorship, and assessment validity. The policy therefore follows a consistent internal pattern in which each clause states what the student must or must not do, explains why

this requirement matters, and sets out what the requirement means in practice across the main assessment types encountered in higher education.

4.1. Critical Thinking and the SAGE Stages

The SAGE framework is grounded in a long-standing scholarly tradition that treats critical thinking as a structured form of reasoned judgement rather than as a generic dispositional virtue. Ennis defined critical thinking as “reasonable, reflective thinking focused on deciding what to believe or do” (Ennis, 2011), a formulation that has remained foundational across subsequent scholarship. The Delphi consensus reported by Facione identified interpretation, analysis, evaluation, inference, explanation, and self-regulation as the core cognitive skills that together constitute critical thinking in the educational context (Facione, 1990). More recently, Bearman and colleagues have argued that the rise of generative AI in higher education places renewed emphasis on evaluative judgement, defined as the capacity to discern the quality of work, including one’s own and that produced by AI tools (Bearman et al., 2024). Evaluative judgement can be understood as the operationalisation of critical thinking within the specific conditions of AI-supported academic work.

Each stage of the SAGE cycle is designed to invite, scaffold, or make visible one or more of these critical thinking dimensions. The Generate stage is intended to protect interpretation by requiring students to articulate a human baseline before engaging with AI, ensuring that the cognitive starting point is the student’s own framing rather than the model’s. The Evaluate stage is intended to operationalise analysis by requiring students to test AI claims against disciplinary anchors and authoritative sources. The Refine stage is intended to engage evaluation and inference, requiring documented accept–modify–reject decisions that record the reasoning behind each revision. The Audit stage is framed as a meta-evaluative move, requiring students to test the validity of AI-generated critique itself rather than accepting it as authoritative. The Reflect stage is intended to correspond to self-regulation, asking students to make explicit their assessment of trust, limits, and remaining uncertainty in the work produced. The Defend stage, finally, is intended to operationalise explanation under scrutiny, requiring the student to articulate the reasoning represented in the final submission in a way that is observable and assessable.

The mapping is therefore a design rationale rather than an empirical finding. SAGE is not equivalent to critical thinking, and not every act of verification or reflection constitutes critical thinking in the full sense intended by Ennis or Facione. The framework’s contribution is to provide a staged scaffold in which the cognitive skills that together constitute critical thinking can be exercised, made visible, and assessed within an AI-permeable assessment environment. The policy advanced in this paper is the institutional expression of that scaffold.

This mapping should not be read as evidence that SAGE reliably produces measurable cognitive gains. The present article does not demonstrate that students who use SAGE improve in interpretation, analysis, evaluation, inference, explanation, or self-regulation. Whether SAGE-AP produces measurable gains in critical thinking, evaluative judgement, epistemic trust calibration, or learning transfer remains an open empirical question for future research.

4.2. Structural Architecture

The policy comprises fifteen clauses, organised into four logical tiers as shown in Figure 2.

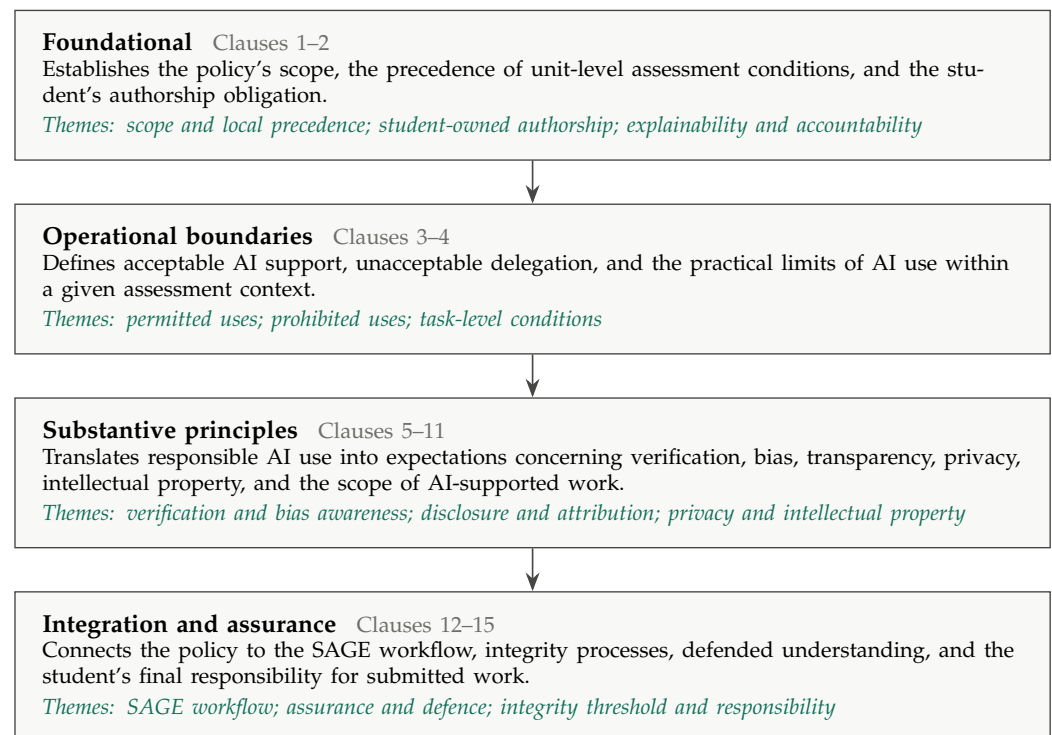


Figure 2. Thematic architecture of SAGE-AP. The figure summarises the four functional tiers, their core policy functions, and the main themes addressed across the fifteen clauses. Full clause-level detail is provided in Table 1.

Table 1 presents a condensed student-facing version of SAGE-AP in its intended policy voice. The complete policy instrument, including the full rule–rationale–practice wording for all fifteen clauses, is provided as Supplementary File S1. The maintained and adaptable version is also available through the SAGE-AP site (Elkhodr & Gide, 2026b).

The tiering reflects a functional logic rather than an explicit numbering convention in the published policy. The foundational tier establishes the policy’s relationship to the institutional environment and the core authorship principle. The operational boundaries tier provides the concrete guidance that most students and educators will seek first, answering the question of what AI may and may not be used for. The substantive principles tier addresses the pedagogical content that distinguishes SAGE-AP from most institutional documents. The integration and assurance tier links the principles back to the SAGE cycle and to the institution’s broader integrity architecture.

Table 1. SAGE-AP: condensed student-facing clause wording.

Policy Tier	Clause	Requirement	Condensed Student-Facing Policy Wording
1. Foundational	1	Scope and unit-level precedence	- Check the course outline, unit profile, assessment brief, or teaching-team instruction before using AI. - Use AI only to the extent permitted for this specific task; do not assume that rules for one assessment apply to another.
	2	You are the author of your work	- Ensure that the final submission reflects your own understanding, analysis, and judgement. - Do not submit work unless you can explain why each important point is included, what source supports it, and how you reached the conclusion.
2. Operational boundaries	3	What you may use AI for	- Use AI, where permitted, to support starting, structuring, checking, critique, language refinement, search discovery, reflection coverage, code assistance, or visual/design development. - In every case, keep the intellectual responsibility, verification, and final judgement with you.
	4	What you must not use AI for	- Do not submit AI-generated work as your own or delegate the core task to AI. - Do not cite unverified AI references, outsource evaluative judgement, fabricate logs or reflection, conceal AI use in group work, enter restricted information, infringe copyright, or submit work you cannot defend.

Table 1. Cont.

Policy Tier	Clause	Requirement	Condensed Student-Facing Policy Wording
3. Substantive principles	5	Verify everything AI tells you	- Verify every important AI-generated claim, reference, statistic, technical assertion, or source suggestion against authoritative sources before relying on it. - Do not cite or use material that you have not personally checked.
	6	Be aware of AI bias	- Assess AI output for cultural, representational, confirmation, authority, gender, racial, disciplinary, and contextual bias. - Do not assume that an AI answer is neutral, current, balanced, or appropriate for your assessment context.
	7	Disclose and attribute AI use correctly	- Disclose AI use in the form required for the task: citation or footnote for quoted or paraphrased AI content, acknowledgement for editing or translation, and structured declaration where process evidence is required. - Remember that attribution is necessary but not sufficient if the intellectual work has been outsourced.
	8	Protect personal information and confidential data	- Do not enter personal, confidential, proprietary, patient, client, commercial-in-confidence, or otherwise restricted information into public AI tools. - Use institutionally approved platforms where available, and do not use AI tools to detect AI-generated content in your own or others' work.
	9	Respect copyright and intellectual property	- Do not paste copyrighted material into AI tools without permission, and do not submit AI output that reproduces protected work. - You are responsible for both what you enter into AI systems and what you include in your submission.
	10	Understand what counts as AI use	- Treat deliberate, task-directed generative AI use as within scope, including AI paraphrasing, Microsoft Copilot or other AI tools in Word and AI code generation where substantial content is produced. - Incidental spell-checking, predictive text, or basic autocomplete is outside scope unless the assessment says otherwise.
4. Integration and assurance	11	These rules apply to all forms of AI output	- Apply the same rules to AI-generated or AI-shaped text, code, images, diagrams, audio, video, data analysis, design prototypes, and other artefacts. - If AI produced or substantially shaped part of the submission, disclose and manage it under this policy.
	12	How the SAGE steps work	- Work through Generate, Evaluate, Refine, Audit, Reflect, and Defend: start from your own understanding, check AI output against authoritative anchors, record accept–modify–reject decisions, judge AI critique independently, and reflect on strengths and limits. - Demonstrate your reasoning under a controlled assurance condition.
	13	Academic integrity	- Treat predominantly AI-generated work, fabricated or back-filled process logs, concealed AI use in group work, and inability to explain submitted work as serious integrity matters. - Understand that these matters may be referred to institutional procedures.
	14	What happens if you cannot fully defend your work	- Distinguish good-faith gaps in understanding from misconduct. - If you engaged honestly but cannot fully defend part of the work, the outcome may be feedback, resubmission, or grade adjustment. - Fabrication, concealment, or deliberate substitution is handled through integrity procedures.
	15	Student responsibility statement	- By submitting the work, you confirm that it reflects your own reasoning and complies with the stated AI conditions. - Confirm that the work has been independently verified, discloses material AI use, protects restricted information, and can be explained and defended by you.

4.3. Four Clauses Illustrating Pedagogical Character

Four clauses warrant closer examination because they mark substantive departures from prevailing institutional practice and show how the broader SAGE research arc is translated into policy form.

Clause 2 (authorship) and the diagnostic test. Box 1 reproduces one clause in full to illustrate the policy's rule–rationale–practice structure and its direct student-facing voice.

Most institutional AI policies address authorship through general statements about originality and honest representation. SAGE-AP operationalises authorship through a diagnostic test that the student can apply before submission. This reframes authorship from a declarative concept to a testable property of the submitted work. It also aligns the policy with the Defend assurance mechanism that operates at the end of the SAGE cycle, in which the same diagnostic is applied under supervised conditions.

Box 1. Clause 2: You are the author of your work.

The purpose of assessment is to demonstrate what you know, what you can do, and how you reason within your discipline. AI is permitted as a support tool, but the final submission must reflect your own understanding, your own analysis, and your own judgement.

Why this matters: your degree certifies that you have developed specific capabilities. If AI performs the intellectual work and you simply present the result, the assessment cannot verify your learning.

What this means in practice:

- In essays, reports, and written assessments, your own argument, synthesis, and critical evaluation must be visible.
- In coding and technical tasks, you must understand, explain, and be able to modify the code you submit.
- In presentations, the substantive argument and spoken explanation must be your own.
- In lab reports and practical work, your own interpretation of data and clinical or design reasoning must drive the submission.
- In research tasks, the identification, reading, evaluation, and synthesis of sources must be performed by you.
- In group assessments, your individual intellectual contribution must remain distinguishable.
- In reflective tasks, the reflective content must be genuinely yours. AI cannot reflect for you.
- In creative and design work, AI-generated outputs must be disclosed and must not be presented as your original creative output unless the assessment explicitly permits this.

The test: if you cannot explain why a particular point is in your submission, what source supports it, or how you arrived at a conclusion, the work does not satisfy the authorship requirement—regardless of how much you edited the AI output.

Clause 4 (non-permitted uses) and the indefensibility criterion. The final prohibited use listed in the clause is:

Producing work you cannot defend: submitting work produced with AI assistance that you are unable to explain, justify, or reproduce under supervised conditions. If you cannot defend it, you do not own it.

This criterion is deliberately placed at the end of the non-permitted uses list because it functions as a backstop for the entire policy. Regardless of whether any other individual prohibition applies, a submission that the student cannot explain under supervised conditions falls outside the policy's permitted scope. This establishes a direct link between the earlier, generative stages of the SAGE cycle, in which AI use is permitted and unrestricted (Steps 1 to 5), and the Defend stage, at which the student's understanding is assured (Step 6). Because SAGE is a single staged process rather than a sorting of work into separate secure and open tracks, open AI-supported engagement during learning and accountability at the point of assurance form one continuous workflow.

Clause 7 (attribution) and the limits of acknowledgement. The clause includes the following statement:

Correct attribution does not automatically resolve an integrity issue. A predominantly AI-generated submission remains problematic even if properly acknowledged, because the intellectual work has been outsourced.

This position marks a distinction from common operational approaches to permitted AI use. In many institutional documents, acknowledgement and disclosure are among the most visible operational mechanisms for managing permitted AI use. However, disclosure does not by itself establish that the submitted work reflects student reasoning, disciplinary judgement, or defensible authorship. What SAGE-AP seeks to avoid is a generic liability-protection declaration, analogous to a "may contain traces of nuts" warning on food labels, that merely signals possible AI involvement without requiring a meaningful account of how AI use shaped the submitted work. SAGE-AP therefore treats correct attribution as necessary but not sufficient. This aligns with the authorship principle in

Clause 2 and with the Defend mechanism operationalised in Clause 12; Clause 14 then specifies what follows when a student cannot fully defend the work. It also addresses a gap in institutional policy where transparency may otherwise be treated as a proxy for legitimate practice.

Clause 12 (Defend) and the controlled assurance moment. A further feature of the policy is that Clause 12 does not treat disclosure, annotation, or process logging as sufficient in themselves. Instead, it links responsible AI-supported work to one controlled point at which the student must explain, justify, or demonstrate the reasoning represented in the submission. In practice, this may take the form of a short viva, a live demonstration, or a timed exercise. The importance of Clause 12 is that it closes the gap between process transparency and demonstrated understanding. Students may produce fluent records of AI use without genuinely owning the reasoning in the final work; the Defend step is designed to test that ownership directly. Clause 14 complements this mechanism by distinguishing competence outcomes from misconduct outcomes when a student cannot fully defend the submitted work. These clauses give the policy an explicit assurance logic that complements approaches relying on disclosure statements, appendices, or retrospective logs (Elkhodr & Gide, 2026b).

A scalable Defend implementation distributes assurance across the cohort and across existing assessment points rather than concentrating it in a single high-stakes event. It therefore need not require a full viva for every student. For example, in a 200-student written assessment, all students could submit a short defence note of 150–200 words identifying one major AI-supported decision, one authoritative source used for verification, and one revision made after evaluating AI output. The teaching team could then conduct a sampled and risk-triggered Defend checkpoint: a random 10–15% sample of the cohort, plus any submissions where authorship, process evidence, or verification appears unclear, is invited to a five-minute oral explanation, live demonstration, or screen-based walkthrough. In a 200-student cohort, a 15% sample would involve 30 short exchanges, or approximately 150 minutes of direct checking, which can be distributed across tutors using a common rubric and calibration examples. The purpose is not to add a single concentrated examination at the end of the task but to distribute assurance proportionately, so that defended understanding is embedded in the ordinary flow of the unit rather than deferred to one high-stakes endpoint. A full viva or final examination remains a legitimate Defend format where a discipline or accreditation requirement calls for it; the objection is to concentrating assurance at a single point, not to examinations themselves. The written defence note also functions as a formative exercise, because the act of articulating one major AI-supported decision and one verification step requires students to rehearse the reasoning that the sampled checkpoint then confirms; even those not selected for the oral exchange therefore benefit from the structured reflection.

4.4. Companion Materials

A policy document in isolation is difficult to adopt. SAGE-AP is therefore accompanied by companion resources intended to support implementation and local adaptation: scaled declaration templates, worked examples, a student-facing FAQ, and a framework-neutral variant for educators who prefer not to invoke the SAGE label. These materials should be read as practical supports rather than validated instruments. Informal use and adaptation across prior SAGE implementation contexts informed their design, but this article does not report a formal pilot evaluation of the companion materials as a policy implementation kit. The complete SAGE-AP instrument is supplied as Supplementary File S1, while the maintained and adaptable policy, declaration templates, examples, FAQ, and neutral

version are available at <https://sage-framework.com/policy> (accessed on 22 May 2026) under a Creative Commons Attribution 4.0 International licence (Elkhodr & Gide, 2026b).

5. Positioning SAGE-AP Relative to Current Institutional Approaches

SAGE-AP is best understood as complementary to, rather than a replacement for, existing institutional policy and assessment frameworks. Two-lane assessment architectures clarify where AI use is restricted or permitted; AI-use scales such as the AIAS help educators align permission levels with learning outcomes; acknowledgement and declaration models improve transparency; and viva-style or demonstration-based mechanisms support assurance of authorship and understanding (City St George's, University of London, 2026; Monash University, 2026; Perkins et al., 2024; University College London, 2026; University of Sydney, 2025). These developments are important because they show that the sector has moved beyond simple prohibition. In practice, the AIAS is set by the coordinator or teaching team to communicate the permitted level of AI use for a given task, whereas SAGE-AP is positioned as the complementary student-facing handout that translates that permitted level into a structured process for responsible engagement during task completion.

The policy-design space addressed by SAGE-AP is narrower and more process-oriented. It brings together human starting points, anchor-based verification, iterative refinement, reflection, process evidence, and defended authorship into a staged student-facing workflow. The contribution is therefore not that SAGE-AP is the only possible response but that it provides one coherent policy instrument for connecting permission settings with the way students actually work during AI-supported assessment.

6. Implications for Higher Education Policy and Practice

Several implications follow from this work. First, universities should avoid treating disclosure as a proxy for pedagogy. A student may disclose AI use accurately and still engage with AI in ways that are superficial, overly dependent, or weakly connected to disciplinary judgement. Disclosure remains important, but it does not by itself demonstrate learning, verification, or authorship.

Second, policy should be expressed in forms that students can actually use while completing the assessment. Broad ethical statements are necessary but often insufficient. Students benefit from staged guidance that clarifies how to proceed, what to document, how to test AI output quality, and what it means to remain accountable for the final submission.

Third, institutions may need to design clearer connections between open AI-supported work and assurance-oriented moments. SAGE-AP does not assume that every task must be fully secured or fully open. It supports a design logic in which students may use AI productively during learning while still being required to defend reasoning, understanding, and final decision-making. In this respect, policy adoption is also an assessment-design task: permission settings, evidence requirements, and Defend checkpoints must be calibrated to task stakes, cohort size, and disciplinary purpose.

Equity and Implementation Considerations

A pedagogically structured policy of the kind advanced here carries implementation costs. The first is the resourcing of the Defend step. Short viva, live demonstration, or timed assurance activities become demanding in large cohorts if applied universally. Institutions adopting SAGE-AP should therefore consider sampled, staged, triggered, or assessment-embedded Defend models, following the worked example in Section 4.4, so that assurance does not become an unmanageable workload for coordinators, early-career academics, or casual teaching staff. Empirical evaluation of SAGE in unsupervised take-home assessment has documented this concern directly: auditing AI-supported

process evidence approximately doubled marking time and shifted educators from assessing learning towards interpreting process logs, a marker's dilemma that proportionate, sampled, and embedded Defend models are intended to contain (Elkhodr & Gide, 2026a).

The second is the documentation burden on students. Graduated declaration templates can calibrate evidence requirements to task stakes, but even light documentation may be experienced unevenly by students for whom English is an additional language, students with cognitive or learning-related disabilities, and students managing significant external commitments. Implementation should therefore include accessible templates, clear exemplars, and tutorial-level support.

The third is differential access to AI tools. Premium systems may offer stronger performance, longer context windows, or more reliable outputs than free tools. A policy that assumes equal AI access risks reproducing digital inequity. Institutions should consider providing approved AI platforms where possible, both to support equity and to provide a safer data-handling environment. Where institutional provision is unavailable, assessment design should avoid privileging students with access to paid tools.

These considerations do not undermine the case for pedagogically structured policy. They show that adoption is itself a learning-design exercise requiring attention to resourcing, accessibility, and equity.

7. Limitations and Conclusions

This paper does not claim to provide a systematic review or prevalence analysis of university GenAI policies or guidance documents. The institutional examples were selected purposively to illustrate visible policy logics across different jurisdictions and assessment contexts. The analysis is therefore interpretive and illustrative rather than representative. The public document corpus is limited to seven institutions and cannot determine what may exist in internal learning management systems, staff-only guidance, local assessment briefs, or teaching practice. The study also did not calculate a formal inter-rater agreement statistic, such as Cohen's kappa or percentage agreement, because the analysis was used to support a bounded conceptual argument rather than to produce a quantitative content-analysis result. This remains a methodological limitation and should be addressed in future, larger-scale policy mapping.

The conceptual contribution is also bounded. SAGE-AP is presented here as a theoretically grounded policy proposal and implementation logic, not as an empirically validated policy intervention. The critical thinking mapping in Section 4 identifies the cognitive work that SAGE is designed to scaffold, but it does not demonstrate that students reliably develop those capabilities by using SAGE-AP. Similarly, the declaration templates, worked examples, frequently asked questions, and framework-neutral policy variants are described as companion implementation supports rather than validated instruments. Future research should extend this work through a larger coded sample of institutional policies and guidance documents, longitudinal analysis of policy change, student and educator usability studies, workload analysis of Defend implementation, and empirical evaluation of whether SAGE-AP improves evaluative judgement, epistemic trust calibration, verification behaviour, and defended authorship.

Higher education policy on generative artificial intelligence has developed significantly in a short period of time. Universities have moved beyond blanket prohibition toward more differentiated approaches that include conditional permission, acknowledgement requirements, documentation expectations, and, in some cases, endpoint assurance mechanisms. These are significant developments. However, the literature and selected public policy examples reviewed in this paper suggest that many institutional approaches

remain more developed in relation to governance, permission, disclosure, and assurance than in relation to student-facing pedagogical processes.

Three practical recommendations follow. First, universities should distinguish explicitly between permission settings and pedagogical process guidance so that students are not left to infer responsible AI practice from compliance statements alone. In practical terms, educators who already use a permission scale such as the AIAS can retain that scale for setting the assessment condition and attach SAGE-AP as the corresponding student-facing process guide, rather than replacing one instrument with the other. Second, educators adopting AI-permitted assessment should use proportionate documentation and assurance requirements, with lighter evidence for low-stakes tasks and stronger process evidence or sampled Defend checkpoints for higher-stakes work. Third, future policy development should be accompanied by empirical evaluation, including student usability studies, educator workload analysis, and investigation of whether structured AI-use policies improve verification practice, evaluative judgement, calibrated trust in AI, and defended authorship.

This perspective has argued that an important missing element in much public-facing policy is a pedagogical middle layer between permission and misconduct. Students need more than boundary setting. They need a structured way to engage with AI responsibly during the process of academic work. The SAGE framework addresses this need by positioning AI-assisted assessment as a staged educational process grounded in human starting points, anchor-based checking, visible judgement, iterative refinement, reflection, and defended authorship. As generative tools increasingly acquire agentic capabilities that can plan and carry out multi-step work with limited human input, a pedagogical layer that keeps verification, evaluative judgement, and defensible authorship with the student is likely to become more important rather than less.

In this sense, SAGE-AP is not simply a policy statement about whether AI is allowed. It is a pedagogically explicit proposal for how AI may be engaged in ways that remain educationally meaningful and academically defensible. As universities continue to refine their assessment policies in the generative AI era, frameworks of this kind may offer an alternative path to models centred solely on permission, prohibition, or post hoc disclosure. Other pedagogical responses to the same problem are also possible; SAGE-AP represents one approach rather than a unique solution.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/educsci16060986/s1>, Supplementary File S1, Full SAGE-AP.

Author Contributions: Conceptualization, M.E. and E.G.; methodology, M.E.; formal analysis, M.E.; investigation, M.E.; resources, M.E. and E.G.; writing—original draft preparation, M.E.; writing—review and editing, M.E. and E.G.; visualization, M.E.; project administration, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work received Open Access article processing charge (APC) funding from Central Queensland University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new empirical data were created or analysed in this study. The article draws on publicly available institutional policy and guidance documents and the publicly available SAGE-AP resource. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5.4 and GPT-5.5) for editorial support, including language refinement, structural review,

condensation of text, figure-caption drafting, and assistance with LaTeX wording. The authors reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors are developers/co-developers of the SAGE framework and maintain or contribute to the SAGE online resources discussed in this article. SAGE-AP is made available under a Creative Commons Attribution 4.0 International licence. No financial conflict of interest is declared.

Appendix A. Document Corpus and Extraction Matrix

Table A1. Scope of the purposive policy document analysis.

Selection Element	Scope	Rationale
Document type	Publicly available university GenAI assessment policies and guidance pages	Ensured that the analysis was based on accessible institutional artefacts rather than internal or anecdotal material.
Jurisdictions	Australia, United Kingdom, Singapore, and China	Provided cross-jurisdictional variation while keeping the analysis bounded for a perspective article.
Institutional examples	University of Sydney; Monash University; University of Technology Sydney; University College London; City St George's, University of London; National University of Singapore; Tsinghua University	Selected to illustrate different policy and guidance emphases, including conditional permission, disclosure, process evidence, critical checking, and defended authorship.
Inclusion criteria	Public, institutionally hosted, assessment-relevant, and directly concerned with student GenAI use	Focused the analysis on documents relevant to assessment policy rather than general AI strategy.
Exclusion criteria	Staff-only pages, media commentary, procurement advice, generic AI strategy, and non-assessment digital transformation material	Reduced conceptual drift and protected the analysis from overclaiming.
Analytic focus	Permission, disclosure, process documentation, verification, privacy and intellectual property, student-facing workflow, and assurance mechanisms	Matched the analysis to the paper's concern with the visibility of student-facing pedagogical process guidance.

Table A2. Document corpus reviewed in the policy scan.

Institution	Jurisdiction	Document Title	Document Type	Accessed	Source URL
University of Sydney	Australia	The Sydney Assessment Framework	Assessment framework	22 May 2026	https://educational-innovation.sydney.edu.au/teaching@sydney/the-sydney-assessment-framework/
Monash University	Australia	AI statements (TeachHQ)	Teaching guidance	22 May 2026	https://www.monash.edu/learning-teaching/TeachHQ/Teaching-practices/artificial-intelligence/AI-statements
University of Technology Sydney	Australia	A guide to using GenAI in specific assessment tasks	Educator guidance	22 May 2026	https://educationexpress.uts.edu.au/blog/2025/01/22/guide-using-genai-specific-assessment-tasks/
University College London	United Kingdom	Three categories of GenAI use in assessment	Assessment guidance	22 May 2026	https://www.ucl.ac.uk/teaching-learning/generative-ai-hub/three-categories-genai-use-assessment
City St George's, University of London	United Kingdom	Use of Generative AI Tools in Doctoral Research	Doctoral research guidance	22 May 2026	https://www.citystgeorges.ac.uk/research/doctoral-college/doctoral-training-resources/use-of-generative-ai-tools-in-doctoral-research
National University of Singapore	Singapore	Academic Integrity Essentials (NUS Libraries)	Library guidance	22 May 2026	https://libguides.nus.edu.sg/new2nus/acadintegrity
Tsinghua University	China	Guiding Principles for the Application of Artificial Intelligence in Education	University-wide guiding principles	22 May 2026	https://www.tsinghua.edu.cn/en/info/1244/14599.htm

Table A3. Populated extraction matrix across seven institutions and seven analytical dimensions.

Institution	D1: Conditional Permission	D2: Disclosure	D3: Process Documentation	D4: Verification	D5: Privacy and IP	D6: Student-Facing Process Guidance	D7: Assurance Mechanism
Sydney	Two-lane model: secure (Lane 1) versus open (Lane 2); AI cannot be banned in open assessments	Where relevant per task design	Not central to the framework	Encouraged through scaffolded use; not operationalised in this document	Not addressed in this document	Limited; directs to a separate AI for Educators resource	Secure lane provides the assurance moment
Monash	Chief Examiner specifies AI use per task; assessment conditions take precedence	Required where AI use is permitted; structured AI statements	Document chatlogs, prompts, iterative drafts, and process steps; periodic screen grabs encouraged	Cross-reference AI-generated summaries against original sources	Monash-managed Copilot recommended for institutional data handling	Developed; categorises permitted uses (search support, text generation, other outputs) with iteration counts and modification records	Disclosure-based; no specific oral or demonstration requirement
UTS	Task-specific can and cannot tables, with rationale linked to learning outcomes	Implied through assessment-level expectations	Not central to this resource	Students required to verify suggested sources and cross-reference summaries	Not the focus of this guidance	Developed at task level; rationales for permitted and prohibited uses and links to support resources	Not specified; task-level discretion
UCL	Three categories: cannot be used, assistive role, integral role	Required for assistive and integral categories	Not central; staff guidance signals where category-level expectations apply	Critical engagement encouraged; staff prohibited from using GenAI detection tools	Data protection requirements; staff must not submit student work to GenAI	Limited on this page; separate student-facing page referenced	Category 1 assessments include in-person examinations, vivas, and Q&A presentations
City St George's	Conditional for doctoral research; GenAI not to be used during the viva voce	Implied through viva discussion of GenAI use	Implied; AI use may be discussed at viva voce	AI must not substitute the candidate's own research and learning	Detection tools prohibited; intellectual property and confidentiality concerns flagged	Doctoral-researcher facing	Viva voce examination is the primary assurance mechanism
NUS	Permitted with critical use; per-instructor decisions	Three tiers: note, methods section, or appendix	Possible via methods section or appendix	Clear emphasis; the third level of described use is evaluation of AI output for accuracy, relevance, objectivity, and completeness	Plagiarism, intellectual property, and copyright treated explicitly	Library-guide format; three-level framework distinguishing wholesale use, intermediate use with human development, and evaluation	Not specified within this document
Tsinghua	Five core principles; AI as auxiliary tool; instructor specifies use per course	Proper disclosure required; copying or mechanical paraphrasing of AI output prohibited	Not specifically detailed	Multi-source verification required; vigilance against hallucinations and cognitive complacency	Sensitive, classified, or unauthorised data prohibited; algorithmic bias and digital divide noted	Within course-level discretion; instructors explain the principles at the start of the semester	Supervisor full-process oversight for theses and dissertations

Appendix B. Framework-Neutral Policy Variant

The framework-neutral variant restates the substance of the fifteen SAGE-AP clauses as fourteen principles, without using the SAGE label or naming the six-step cycle. It is reproduced here as a quick, brand-independent reference; the full instrument, companion templates, and frequently asked questions are provided in Supplementary File S1 and maintained at <https://sage-framework.com/policy> (accessed on 22 May 2026).

Table A4. Framework-neutral variant of SAGE-AP: fourteen principles for AI-supported assessment.

Principle	What It Requires
1 Check what is permitted	Confirm whether and how much AI use is allowed for this specific task before you begin. Rules for one assessment do not carry over to another.
2 You are the author	The final submission must reflect your own understanding, analysis, and judgement. If you cannot explain why a point is included, what supports it, and how you reached it, the work does not meet this requirement.
3 Permitted uses	Where permitted, you may use AI to start, structure, check, critiquing, and refine, provided verification and final judgement remain with you.
4 Prohibited uses	Do not submit AI-generated work as your own, delegate the core task, cite unverified AI references, outsource judgement, fabricate process evidence, or conceal AI use.
5 Verify everything	Check every important AI-generated claim, reference, or statistic against authoritative sources. Do not cite anything you have not personally checked.
6 Watch for bias	Assess AI output for cultural, confirmation, and authority bias. Do not assume it is neutral, current, balanced, or appropriate to your context.
7 Disclose and attribute	Disclose AI use in the required form. Disclosure alone does not make a predominantly AI-generated submission acceptable.
8 Protect data	Do not enter personal, confidential, or restricted information into public AI tools. Use approved platforms where available.
9 Respect copyright	Do not enter copyrighted material into AI tools without permission, and do not submit output that reproduces protected work.
10 Know what counts	Deliberate, task-directed AI use is in scope, including AI paraphrasing and code generation. Incidental spell-checking and autocomplete are not, unless the assessment says otherwise.
11 All output types	These rules apply equally to AI-generated or AI-shaped text, code, images, audio, data analysis, and design.
12 Work and document	Work from your own understanding, then check AI output, record what you accepted, modified, or rejected and why, and note where AI helped and where it failed.
13 Defensible authorship	Be ready to explain, justify, or demonstrate the reasoning in your submission under supervised conditions at any time. If you cannot defend it, you do not own it.
14 Integrity and good-faith gaps	Predominantly AI-generated work, fabricated evidence, concealed AI use, and inability to explain your work are integrity matters. A good-faith gap in understanding is treated as a competence matter, handled differently from misconduct.

Student declaration. By submitting this work, I confirm that it reflects my own reasoning, complies with the stated AI conditions, has been independently verified, discloses any material AI use, contains no restricted information entered into public AI tools, and can be explained and defended by me.

References

- An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22, 10. [CrossRef]
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjaw, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49, 893–905. [CrossRef]
- Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16, 296. [CrossRef]
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9, 27–40. [CrossRef]
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20, 38. [CrossRef]
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. [CrossRef]
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, 39, 1727–1739. [CrossRef]
- City St George's, University of London. (2026). *Use of generative AI tools in doctoral research*. Available online: <https://www.citystgeorges.ac.uk/research/doctoral-college/doctoral-training-resources/use-of-generative-ai-tools-in-doctoral-research> (accessed on 22 May 2026).
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967. [CrossRef]
- Elkhodr, M., & Gide, E. (2025). The SAGE framework for developing critical thinking and responsible generative AI use in cybersecurity education. *Discover Education*, 4, 517. [CrossRef]
- Elkhodr, M., & Gide, E. (2026a). Embedding assurance within learning: Empirical evidence from the SAGE framework for repositioning take-home assessment in AI-integrated higher education. *STEM Education*, 6, 584–605. [CrossRef]

- Elkhodr, M., & Gide, E. (2026b). *Structured AI-guided education assessment policy (SAGE-AP) for AI-integrated assessment*. Available online: <https://sage-framework.com/policy> (accessed on 22 May 2026).
- Elkhodr, M., & Gide, E. (2026c). *The SAGE framework: Structured AI-guided education*. Zenodo. [CrossRef]
- Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. *Inquiry: Critical Thinking Across the Disciplines*, 26, 4–18. [CrossRef]
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The California Academic Press.
- Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, 14048. [CrossRef] [PubMed]
- Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, 13, 15310. [CrossRef] [PubMed]
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26, 91–108. [CrossRef] [PubMed]
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10, 18–26. [CrossRef]
- Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56, 2522–2549. [CrossRef]
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49, 651–664. [CrossRef]
- McDonald, N., Johri, A., Ali, A., & Hingle Collier, A. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*, 3, 100121. [CrossRef]
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. [CrossRef] [PubMed]
- Monash University. (2026). *AI statements*. Available online: <https://www.monash.edu/learning-teaching/TeachHQ/Teaching-practices/artificial-intelligence/AI-statements> (accessed on 22 May 2026).
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world’s top-ranking universities. *Computers and Education Open*, 5, 100151. [CrossRef]
- National University of Singapore Libraries. (2026). *Academic integrity essentials*. Available online: <https://libguides.nus.edu.sg/new2nus/acadintegrity> (accessed on 22 May 2026).
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. [CrossRef]
- Parker, L., Loper, A. J., Hayes, J., Karakas, A., White, S., & Hallman, H. (2025). Comparative analysis of artificial intelligence policies in universities across five countries. *Discover Computing*, 28, 267. [CrossRef]
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21 (6), 1–18. [CrossRef]
- Perkins, M., & Roe, J. (2024). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, 37, 633–653. [CrossRef]
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. [CrossRef]
- Tertiary Education Quality and Standards Agency. (2023). *Assessment reform for the age of artificial intelligence*. TEQSA. Available online: <https://www.teqsa.gov.au/about-us/news-and-events/latest-news/assessment-reform-age-artificial-intelligence> (accessed on 25 May 2026).
- Tertiary Education Quality and Standards Agency. (2025). *Enacting assessment reform in a time of artificial intelligence*. TEQSA. Available online: <https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/enacting-assessment-reform-time-artificial-intelligence> (accessed on 22 May 2026).
- Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review*, 4, 356–367. [CrossRef]
- Tsinghua University. (2025). *Tsinghua university releases comprehensive guiding principles for AI use in education*. Available online: <https://www.tsinghua.edu.cn/en/info/1244/14599.htm> (accessed on 22 May 2026).
- University College London. (2026). *Three categories of GenAI use in assessment*. Available online: <https://www.ucl.ac.uk/teaching-learning/generative-ai-hub/three-categories-genai-use-assessment> (accessed on 22 May 2026).
- University of Sydney. (2025). *The Sydney assessment framework*. Available online: <https://educational-innovation.sydney.edu.au/teaching@sydney/the-sydney-assessment-framework/> (accessed on 22 May 2026).

University of Technology Sydney. (2025). *A guide to using GenAI in specific assessment tasks*. Available online: <https://educationexpress.uts.edu.au/blog/2025/01/22/guide-using-genai-specific-assessment-tasks/> (accessed on 22 May 2026).

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, 40. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.