The AI Academic Integrity Toolkit: Designing Ethical and Reflective Assessment Practices in the Age of AI

Rahme Sadikoglu, Anglia Ruskin University London, United Kingdom

The European Conference on Education 2026
Official Conference Proceedings

Abstract

This paper presents the AI Academic Integrity Toolkit, an open educational resource designed to support lecturers in embedding ethical and reflective AI use within assessment design and feedback. Building on the AI-Enabled Academic Integrity (AIAI) Framework (Visin, 2025), the Toolkit translates its four pillars, Ethical Access, Critical Competence, Reflexive Transparency and Educational Value, into assessment-oriented design principles. The Toolkit was developed through reflective engagement with academic integrity review processes, review of institutional policy guidance and ongoing professional discussions with educators. The Toolkit provides adaptable assessment models, suggested assessment brief language, reflection prompts and feedback resources aligned with principles of constructive alignment (Biggs & Tang, 2011) and an educative understanding of academic integrity (Bretag, 2013). Designed for potential application across undergraduate and postgraduate contexts, it aims to clarify boundaries around responsible AI use, promote student reflection and support educators in evaluating AI-assisted work. The Toolkit is presented as a design-based contribution intended to support pedagogical implementation and future evaluation, rather than as a report of empirical outcomes. By reframing integrity as an educational rather than a compliance-based practice, the Toolkit contributes an adaptable, theoretically grounded approach to integrating AI literacy and ethical reflection within higher-education assessment practices.

Keywords: academic integrity, AI literacy, assessment design, higher education, reflective pedagogy

Introduction

Generative artificial intelligence (GenAI) has unsettled established assumptions about the relationship between authorship, originality and learning in higher education. Students can now use AI systems to generate ideas, organise arguments, produce written content and revise existing work, making it increasingly difficult to regard the final submission as a straightforward representation of individual cognitive effort. These developments do not render authorship or originality irrelevant. They do, however, complicate the boundaries between legitimate assistance, collaborative knowledge production and inappropriate outsourcing, and raise important questions about what assessment is intended to evidence in AI-mediated learning environments (Cotton et al., 2024; Fawns, 2019; Kofinas et al., 2025).

Understandably, a prominent strand of the institutional and academic response has focused on academic misconduct, policy revision and the detection of AI-generated text (Cotton et al., 2024; Kofinas et al., 2025). Such responses remain necessary where AI is used deceptively or in ways that prevent students from demonstrating the required learning outcomes. However, detection alone cannot establish how a student used AI, whether its outputs were questioned, what decisions were made during the process or whether the interaction contributed to learning. Academic integrity should therefore not be reduced to the absence of misconduct. As Bretag (2013) argues, integrity also needs to be understood as an educative and developmental concern embedded within curriculum and institutional culture.

The educational challenge is consequently not only to regulate AI use, but to create assessment conditions in which students can learn to use it responsibly. The AI-Enabled Academic Integrity (AIAI) Framework developed by Visin (2025) addresses this challenge through four interconnected pillars: Ethical Access, Critical Competence, Reflexive Transparency and Educational Value. Together, these principles shift attention from the binary question of whether AI was used towards the quality, purpose and transparency of that use. Yet conceptual principles do not automatically translate into everyday assessment practice. What remains less developed is how lecturers can embed such expectations within assessment briefs, reflection activities, marking criteria and feedback without requiring extensive changes to existing modules.

This paper responds to that practical gap by presenting the AI Academic Integrity Toolkit (Sadikoglu, 2025). Building on the AIAI Framework (Visin, 2025), the Toolkit operationalises its four pillars through adaptable resources for assessment design, student reflection and feedback. Informed by constructive alignment, it proceeds from the position that ethical and critical AI use must be made visible within the assessment process if it is to become a meaningful part of student learning (Biggs & Tang, 2011). I argue that the Toolkit offers a practical route for moving academic integrity from a predominantly reactive concern towards a more educative practice, while retaining the need for clear institutional boundaries. Before presenting the Toolkit in detail, the following section situates this contribution within debates on academic integrity, GenAI and constructive alignment. The paper then explains the practice-informed design process, introduces the Additive, Embedded and Developmental integration models, and considers their intended applications and limitations.

From Detection to Educational Integrity: Literature and Conceptual Foundations

The International Center for Academic Integrity (2021) articulates academic integrity through six fundamental values: honesty, trust, fairness, respect, responsibility and courage. These values remain fundamental to higher education, but their institutional application has often centred on rule compliance, individual authorship and the prevention of misconduct. Bretag (2013) argues that institutions should move beyond deterrence, detection and punishment by embedding academic integrity within assessment, curriculum and institutional culture. This educative position is particularly important in the context of generative artificial intelligence, where the distinction between legitimate academic support and inappropriate substitution of student effort has become increasingly difficult to establish through the final submitted product alone.

GenAI systems can support students in generating ideas, structuring arguments, summarising information and producing written content. Their presence within educational practice therefore makes it more difficult to infer a student's individual contribution and learning from the submitted text alone (Cotton et al., 2024; Rudolph et al., 2023). This does not mean that individual responsibility or authentic learning have ceased to matter. Rather, it means that authorship is increasingly shaped by interactions between the student, the assessment task and the technological tools available to them. Fawns's (2019) account of postdigital education is useful here because it rejects the idea that digital technologies function merely as external additions to otherwise unchanged educational practices. Instead, technologies form part of the conditions through which learning, writing and knowledge production take place. The integration of AI should not, however, be assumed to produce educational value in itself. O'Dea and O'Dea (2023) observe that evidence concerning the pedagogical impact of AI in higher education remains limited, despite widespread claims about its potential. This reinforces the need to examine what a particular use contributes to learning rather than equating technological adoption with educational benefit.

These developments expose the limitations of responses that concentrate primarily on detecting whether AI has been used. Detection and regulation remain necessary where students conceal AI use, breach clearly stated assessment requirements or submit work that does not demonstrate the intended learning outcomes. However, the identification of possible AI-generated content cannot, by itself, explain how a student worked with AI, why particular outputs were accepted or rejected, or whether the interaction supported understanding. Kofinas et al. (2025) found that markers generally could not distinguish assessments with GenAI input from those without it and that authentic assessment alone did not safeguard academic integrity. The educational question is consequently broader than whether AI was present within the process. It also concerns the quality of the student's judgement, the transparency of their decisions and the relationship between AI use and the purpose of the assessment.

I therefore approach academic integrity not only as a regulatory responsibility but also as an educational practice. From this perspective, students should be supported in developing the capacity to make informed decisions about when AI use is appropriate, to question the reliability and limitations of generated outputs, and to explain how those outputs were incorporated into their work. Integrity is not replaced by AI literacy within this formulation. Rather, critical and ethical AI literacy becomes one of the contemporary conditions through which integrity is enacted. This position retains the importance of institutional boundaries while recognising that students require more than prohibitions if they are to navigate AI-mediated learning responsibly.

Moving beyond a final-product focus requires these expectations to be made visible before and during assessment. This is particularly important because assessment influences what students attend to, how they distribute their effort and how they approach learning (Gibbs & Simpson, 2005). Constructive alignment provides a basis for connecting these expectations with intended learning outcomes. Biggs and Tang (2011) argue that learning activities and assessment tasks should be aligned with those outcomes. If ethical judgement, critical evaluation and transparency are valued educational outcomes, they must therefore be made visible within assessment design rather than assumed or addressed only after suspected misconduct has occurred. Assessment briefs communicate what students are expected to do, marking criteria indicate what will be valued, and feedback helps students understand the quality of their current performance and how it might be developed (Carless & Boud, 2018; Price et al., 2012). In the context of GenAI, constructive alignment means connecting guidance on AI use with the purpose of the task, the required evidence of learning and the feedback students receive.

The AI-Enabled Academic Integrity Framework (Visin, 2025) offers a conceptual structure for this alignment. Its first pillar, Ethical Access, concerns the student's judgement about when and why AI use is appropriate within a particular educational context. Critical Competence refers to the capacity to construct purposeful prompts and evaluate AI outputs for accuracy, bias, relevance and disciplinary suitability. Reflexive Transparency extends beyond a basic declaration by asking students to explain which tools were used, how outputs were selected or modified, and why particular decisions were taken. Finally, Educational Value considers whether AI contributed meaningfully to understanding or skill development rather than functioning simply as a shortcut to task completion. The four pillars are interconnected: appropriate access without critical evaluation is insufficient, while transparency has limited educational value if it does not reveal purposeful engagement with learning.

An established practical response to these challenges is the Artificial Intelligence Assessment Scale (AIAS), which enables educators to specify and communicate the permitted role of GenAI within an assessment and has subsequently been developed as a framework for task redesign and assessment validity (Perkins, Furze, et al., 2024; Perkins, Roe, et al., 2025). The AI Academic Integrity Toolkit is complementary to this work rather than a substitute for it. Whereas the AIAS establishes the role and scope of permitted AI use within a task, the Toolkit focuses on how students explain and evaluate their decisions, how that evidence can inform assessment criteria, and how feedback can support further development.

The AIAI Framework consequently offers a basis for moving from the binary distinction between AI use and non-use towards a more considered evaluation of purpose, process and judgement. What remains less developed, however, is how these principles can be incorporated into the everyday structures of assessment without requiring complete curriculum redesign. The AI Academic Integrity Toolkit responds to this practical gap by operationalising the four pillars through adaptable assessment models, student reflection and feedback practices (Sadikoglu, 2025). The following section explains the practice-informed design process through which these resources were developed.

Developing the AI Academic Integrity Toolkit: A Practice-Informed Design Approach

As established in the previous section, the central difficulty is not simply the absence of ethical principles, but the lack of practical routes through which they can be embedded within assessment. The AI Academic Integrity Toolkit was developed as a practice-informed response to this problem (Sadikoglu, 2025). The Toolkit was therefore not conceived as a separate policy document, nor as a replacement for existing academic integrity procedures. Instead, it was designed to help educators incorporate ethical and critical AI use into the ordinary structures through which assessment expectations are communicated and learning is evaluated.

Its development was informed by four interconnected sources. First, reflective engagement with academic integrity review processes drew attention to an evidential gap that arises when attention is directed primarily towards the final submission. Identifying which tools were used does not necessarily reveal why particular outputs were accepted, rejected or modified, or how those decisions related to the intended learning outcomes. Second, institutional guidance provided an important regulatory context by clarifying the distinction between permitted educational support and forms of AI use that may prevent authentic demonstration of learning. Third, scholarship on constructive alignment, assessment literacy, feedback and academic integrity helped position these practical concerns within a wider pedagogical framework (Biggs & Tang, 2011; Bretag, 2013; Carless & Boud, 2018; Price et al., 2012). Finally, ongoing discussions with educators and reflection on my wider academic support practice helped shape the resource by drawing attention to disciplinary variation, validated assessment structures, staff workload and the need for accessible, discipline-sensitive student-facing language supported by concrete examples.

I describe this as a practice-informed design process rather than an empirical study. The professional discussions that informed the Toolkit were not used as formally collected research data, and no claims are made here about educator or student outcomes. No student case records, identifiable case material or formally collected case data were analysed for this paper. Their role was developmental: they helped identify practical questions that the Toolkit needed to address and informed decisions about its structure, flexibility and language.

Several design principles followed from this process. The Toolkit needed to be adaptable across disciplines, proportionate in relation to staff workload and sufficiently flexible to operate within different institutional and assessment contexts. It also needed to retain a developmental orientation. Students should not be expected merely to declare whether AI was used; they should be supported in explaining the reasoning behind that decision, how any AI-generated outputs were evaluated and what educational value, if any, AI use produced. For this reason, the Toolkit connects assessment design, reflection and feedback rather than approaching hem as separate activities. Its purpose is not to impose a single institutional model, but to provide educators with a set of adaptable pathways through which the four AIAI pillars can be made visible in assessment practice.

The AI Academic Integrity Toolkit: Structure and Assessment Applications

This section turns to the Toolkit itself and examines how the four pillars of Visin's (2025) AIAI Framework are translated into practical resources for assessment design, student reflection and feedback (Sadikoglu, 2025). The Toolkit is organised around the principle that expectations concerning AI use should be communicated before students complete an assessment, considered during the learning process and revisited through feedback afterwards. In this sense, academic integrity is positioned not as a separate intervention that begins only when misconduct is suspected, but as part of the ordinary design of assessment.

The Toolkit contains three connected areas of practice. The first concerns the design and communication of assessment. This includes three integration models through which lecturers may incorporate the AIAI principles into existing assessment structures, together with suggested language for assessment briefs. The second concerns student reflection. A self-assessment form invites students to explain whether AI was used and the reasoning behind that decision. Where AI was used, students explain how they constructed prompts, evaluated the outputs produced and decided what to accept, reject or modify. The form also asks them to identify where AI contributed to the assessment and to consider whether its use supported their learning. The third concerns feedback and feedforward. The Toolkit provides traffic-light feedback, short pillar-based comments and reflective questions intended to help students identify where their engagement was strong and where further development may be needed (Sadikoglu, 2025).

These components are intended to operate together. Assessment briefs communicate what is expected, reflection makes the student's process more visible, and feedback provides an opportunity to develop future judgement. This sequence reflects the principles of constructive alignment because it connects what students are asked to do with what educators subsequently evaluate and discuss with them (Biggs & Tang, 2011). It also draws on assessment literacy, which emphasises that students need access to the language, standards and expectations through which academic work is judged (Price et al., 2012). In the context of GenAI, this means that students should not be expected to infer the boundaries of acceptable use or understand automatically what critical and transparent AI use involves.

Table 1 summarises the three integration models, which are discussed in more detail below. Together, they provide different routes for embedding the AIAI principles within assessment practice.

Table 1
Three Models for Integrating AIAI Into Assessment Practice

Model	Approach	Potentially suitable for
Additive	Introduces AI use and academic integrity as a separate criterion within the marking scheme, allowing the four AIAI pillars to be assessed directly.	Assessments in which AI use is an explicit requirement and there is sufficient flexibility to add or revise a marking criterion.
Embedded	Incorporates relevant AIAI principles within existing assessment or marking criteria, with evidence from the self-assessment form informing how those criteria are applied.	Modules with validated or relatively fixed marking schemes, or where the AIAI principles align naturally with existing criteria.
Developmental	Uses the AIAI self-assessment as a non-graded reflective activity, supported by formative feedback on the four pillars.	Foundation, early-stage or skills-development modules where the priority is developing awareness and reflective practice.

Note. Adapted from *A Practical Toolkit for Ethical and Reflective Assessment Design in the Age of AI* (Sadikoglu, 2025).

The Additive model introduces AI use and academic integrity as a distinct assessment or marking criterion. For example, a module may include a dedicated criterion that evaluates how appropriately, critically and transparently a student used AI, the decisions they made during the process and whether that use supported the intended learning outcomes. This approach makes expectations highly visible and allows the four AIAI pillars to be assessed directly rather than remaining implicit within broader criteria. It may therefore be particularly appropriate where ethical and reflective AI use forms an explicit part of the assessment. Introducing a separate criterion may, however, require formal approval or revision of existing marking arrangements, depending on institutional and programme-level processes. The wording and weighting of the criterion should also remain proportionate to the wider assessment task so that AI use does not overshadow the disciplinary knowledge and skills being assessed. For this reason, the Additive model is most appropriate where AI use is an explicit assessment requirement rather than simply an optional form of support. The assessment brief should clearly specify the permitted scope of AI use, the evidence students are expected to provide and how their use will be evaluated. Where AI use remains optional, the Embedded or Developmental model may offer a fairer and more appropriate route.

The Embedded model incorporates the AIAI principles within criteria that already exist. A criterion concerning critical analysis, for instance, may include consideration of how a student evaluated the accuracy, relevance or bias of AI-generated material, while a criterion concerning reflective practice may include their explanation of why AI was used, which outputs were accepted or rejected and how those outputs were revised. Evidence for these judgements can be provided through the AIAI self-assessment form, submitted alongside the main assessment. The marker can then consider the student's responses when applying the relevant existing criteria, rather than awarding a separate mark for AI use. Where AI use is optional, choosing not to use it should not in itself reduce the mark; the existing criterion should continue to assess the academic capability it was designed to measure. This model may be more appropriate where adding a new criterion is impractical or where the AIAI principles align naturally with existing learning outcomes and marking criteria. Its advantage lies in connecting ethical and critical AI use with the academic capabilities that the assessment already values. At the same time, there is a risk that expectations may remain too implicit. The assessment brief should therefore identify which existing criteria relate to AI use, explain how the self-assessment form will inform marking and clarify what students are expected to disclose and reflect upon.

The Developmental model adds a non-graded reflective component to the assessment process. Students complete the AIAI self-assessment and receive formative feedback on the four pillars, without their responses contributing directly to the final mark. Unlike the Additive and Embedded models, the self-assessment is not used as evidence when applying marking criteria. This model may be particularly appropriate for foundation, early-stage or skills-development modules where the immediate priority is to build confidence and reflective habits. It is also intended to create conditions in which students can discuss their AI use more openly, because disclosure is framed as an opportunity for development rather than as evidence of misconduct. Nevertheless, reflection should not be assumed to be straightforward. Students may require examples, simplified language and formative guidance to move beyond brief declarations towards meaningful explanation. A developmental approach does not, however, override institutional academic integrity procedures. A self-assessment may indicate AI use that falls outside the boundaries stated in the assessment brief, and the Toolkit cannot replace the academic integrity procedures that may apply in such circumstances. Its role is instead to make expectations clear from the outset by stating what forms of AI use are permitted, explaining the purpose of the self-assessment and positioning accurate disclosure as part of responsible academic practice. This clarity is important for sustaining trust while avoiding the suggestion that the developmental model operates separately from institutional requirements.

Across the three models, the AIAI self-assessment provides a common structure for making students' decisions more visible, although its function differs according to the model selected. In the Additive and Embedded models, it may provide evidence that informs the application of the relevant criterion or criteria; in the Developmental model, it is used only for reflection and formative feedback. Students are asked to explain whether AI was used and the reasoning behind that decision, describe how prompts and generated outputs were evaluated, identify the tools used and where they contributed to the assessment, and reflect on what, if anything, AI use contributed to their learning. The form therefore moves beyond simple disclosure by making the student's decisions available for reflection, assessment and feedback, as appropriate to the model being used (Sadikoglu, 2025). Table 2 provides an illustrative Embedded-model application, showing how assessment brief language and the self-assessment form can work together.

Table 2
Illustrative Embedded-Model Application

Toolkit Element	**Illustrative wording**
Assessment brief	You may use generative AI tools to generate initial ideas, clarify concepts or obtain feedback on a draft. You may not submit AI-generated text as your own writing or use AI in ways that replace the learning outcomes being assessed. You must submit the AIAI self-assessment form alongside your work. Your responses will inform how the existing criteria for critical analysis and reflective practice are applied; they will not receive a separate mark. Choosing not to use AI will not disadvantage you, provided that decision is explained.
Self-assessment prompts	Ethical Access: Did you use AI in preparing this assessment? Explain the reasoning behind your decision in relation to the purpose of the task. Critical Competence: Where AI was used, how did you evaluate its outputs? What did you accept, reject or modify, and why? Reflexive Transparency: Which tools did you use, and where did they contribute to the submitted work? Educational Value: What, if anything, did AI use contribute to your understanding or skill development?

Note. Adapted from *A Practical Toolkit for Ethical and Reflective Assessment Design in the Age of AI* (Sadikoglu, 2025).

The feedback resources follow the same pillar-based structure. Traffic-light feedback offers a concise indication of whether the student's reflection demonstrates each pillar clearly, partially or not yet sufficiently. Under Critical Competence, for example, green indicates a clear evaluation of the strengths and limitations of specific outputs; amber indicates some evaluation but limited depth; and red indicates that outputs were accepted without critical examination. Short reflective comments allow lecturers to identify specific strengths and areas for development, while feedforward questions encourage students to consider how their future decisions and practices might be improved. This future-oriented emphasis is important because feedback does not automatically influence later practice; students need guidance that is sufficiently specific and actionable to inform subsequent decisions (Carless & Boud, 2018; Gibbs & Simpson, 2005). The Toolkit does not assume that every lecturer will use all three feedback formats. Rather, the resources are intended to be selected and adapted according to module purpose, student level and staff capacity.

Taken together, these components show that the Toolkit offers three adaptable pathways rather than a single fixed model. Legitimate uses of AI vary across disciplines, while institutional policies and assessment regulations also differ. Students may also vary in their familiarity with reflective writing and their ability to explain their decisions. It is important, therefore, that the four AIAI pillars remain consistent while their practical application can be adapted to different contexts. The purpose is not to standardise AI use across disciplines or institutions, but to provide a structure through which expectations can be made clearer and students' decisions can become a more visible part of the assessment process.

Discussion, Limitations and Future Evaluation

The AI Academic Integrity Toolkit offers a set of adaptable pathways for translating ethical principles into assessment design. Its contribution lies not in proposing that AI use should always be encouraged, but in providing structures through which its appropriateness, purpose and educational value can be considered more explicitly. This distinction is important. The presence of AI within an assessment does not, by itself, demonstrate either meaningful learning or academic misconduct. What requires attention is how students use the technology, how their decisions relate to the learning outcomes and whether the assessment continues to provide credible evidence of their knowledge and judgement (Cotton et al., 2024; Kofinas et al., 2025). The Toolkit may therefore support a more educational approach to academic integrity by connecting expectations, reflection and feedback. By aligning expectations concerning AI use with assessment activities, criteria and feedback, it draws together the pedagogical principles discussed earlier in the paper (Biggs & Tang, 2011; Carless & Boud, 2018; Price et al., 2012). This complementarity also has a practical implication. Where an institution already uses the AIAS or another permission-based framework, the Toolkit can operate within the permitted level of use by structuring the reflection, evidence and feedback that follow (Perkins et al., 2025). AI use is consequently positioned as something that can be examined and developed throughout the assessment process rather than considered only after a possible breach has occurred.

Adaptability should not be confused with universal applicability. Disciplines differ considerably in what constitutes legitimate AI assistance, and institutional policies may restrict how assessment briefs or marking criteria can be modified. The Additive model is most appropriate where AI use forms an explicit assessment requirement and therefore depends on clear task instructions, equitable access to the relevant technology and proportionate weighting within the marking scheme. Where AI use is required, institutions should provide access to an approved tool or to tools offering sufficiently comparable functions. Assessment design should not advantage students who can purchase enhanced versions, and guidance should distinguish permitted language or accessibility support from uses that replace the learning being assessed. In the Embedded model, the relationship between the AIAI principles and existing criteria may remain insufficiently visible unless it is stated clearly within the assessment brief and supported by the self-assessment form. There may also be variation in how educators interpret and apply the relevant criteria. The Developmental model, meanwhile, depends on students receiving sufficient guidance and formative feedback to reflect meaningfully on their own practices. The language of the four pillars may also require further scaffolding for students who are unfamiliar with reflective writing or abstract academic terminology (Sadikoglu, 2025).

There are further limitations to the present paper. The Toolkit is introduced as a conceptual and design-based contribution, and no claims are made about its effectiveness. Its influence on student understanding, transparency, academic behaviour or learning has not been examined here. The self-assessment form also relies on students' accounts of their own AI use and should not be regarded as an independent method of verifying how work was produced or whether academic misconduct has occurred. Evidence from mandatory AI-use declarations suggests that fear of academic repercussions, ambiguous guidance and inconsistent enforcement may discourage accurate disclosure (Gonsalves, 2025). The form should therefore support reflection and transparency, but should not be considered as a complete evidential record. Similarly, the practicality of the feedback resources may vary according to class size, marking workload and staff confidence. Where the four pillars contribute to formal marking through the Additive or

Embedded models, consistency of interpretation, moderation and equitable application across educators also remain open questions.

The limitations discussed above provide the basis for future evaluation, which should examine not only whether the Toolkit can be implemented, but how it is understood and adapted in practice. Relevant questions include whether students interpret the four pillars consistently, whether the reflection prompts generate substantive rather than formulaic responses, and how educators distinguish between thoughtful reflection and simple disclosure. Evaluation should also examine how consistently the Additive and Embedded models are applied, whether the Developmental model supports more meaningful reflection over time and how students experience the different uses of the self-assessment form. Research should further consider the manageability of the feedback approaches, equitable access where AI use is required, how justified non-use is addressed in contexts where AI remains optional, and the adaptations required across levels and disciplines. A process of implementation, review and revision would allow these issues to inform the further development of the Toolkit, rather than assuming that it represents a finished solution (McKenney & Reeves, 2012). In this sense, the Toolkit provides a foundation for further pedagogical inquiry, while its educational value remains a matter for empirical examination.

Conclusion

This paper has argued that academic integrity in AI-mediated assessment cannot be addressed solely by determining whether AI was used. The more important educational task is to make the purpose, quality and consequences of that use visible within the assessment process. Building on Visin’s (2025) AIAI Framework, the AI Academic Integrity Toolkit translates Ethical Access, Critical Competence, Reflexive Transparency and Educational Value into three adaptable routes: Additive, Embedded and Developmental. These routes are supported by assessment brief language, student self-assessment and feedback resources designed to help educators connect expectations concerning AI use with the learning outcomes students are expected to demonstrate (Sadikoglu, 2025). The Toolkit does not treat AI use as a universal assessment requirement and recognises that, where it remains optional, students may make a reasoned decision not to use it. Its contribution is to provide a clear structure through which students can justify their choices, evaluate AI-generated outputs and reflect on whether technological assistance has supported genuine learning. Whether these aims are realised must be examined through implementation and empirical evaluation. Nevertheless, by locating integrity within the design of assessment rather than leaving it primarily to post-submission detection, the Toolkit offers a practical foundation for a more transparent, critical and educational response to generative AI in higher education.

Acknowledgements

The author would like to thank colleagues at Anglia Ruskin University London for the professional discussions that informed the development of the AI Academic Integrity Toolkit.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that Grammarly, an AI-assisted writing software, was used in proofreading and refining the language used in the manuscript. The usage included correcting grammatical and spelling errors and rephrasing statements for accuracy, clarity and readability. The author further declares that no other AI or AI-assisted technologies have been used in writing the manuscript. The Toolkit design, selection and interpretation of the literature, analysis and discussion presented in the manuscript are the author's own.

References

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). McGraw-Hill Education.

Bretag, T. (2013). Challenges in addressing plagiarism in education. *PLOS Medicine*, 10(12), e1001574. https://doi.org/10.1371/journal.pmed.1001574

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education,* 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International,* 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Fawns, T. (2019). Postdigital education in design and practice. *Postdigital Science and Education*, 1(1), 132–145. https://doi.org/10.1007/s42438-018-0021-8

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31.

Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education,* 50(4), 592–606. https://doi.org/10.1080/02602938.2024.2415654

International Center for Academic Integrity. (2021). *The fundamental values of academic integrity* (3rd ed.). https://academicintegrity.org/aws/ICAI/asset_manager/get_file/911282?ver=1

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(6), 2522–2549. https://doi.org/10.1111/bjet.13585

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research.* Routledge.

O'Dea, X., & O'Dea, M. (2023). Is artificial intelligence really the next big thing in learning and teaching in higher education? A conceptual paper. *Journal of University Teaching and Learning Practice*, 20(5), 1–17. https://doi.org/10.53761/1.20.5.05

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), 49–66. https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., & Furze, L. (2025). Reimagining the artificial intelligence assessment scale: A refined framework for educational assessment. *Journal of University Teaching and Learning Practice*, 22(7), 1–26. https://doi.org/10.53761/rrm4y757

Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). Assessment literacy: The foundation for improving student learning. Oxford Centre for Staff and Learning Development.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching,* 6(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Sadikoglu, R. (2025). A *practical toolkit for ethical and reflective assessment design in the age of AI* [Preprint]. figshare. https://doi.org/10.6084/m9.figshare.30340876

Visin, G. (2025). *AIAI – AI-enabled academic integrity* [Preprint]. figshare. https://doi.org/10.6084/m9.figshare.28827620.v1

Contact email: rahme.sadikoglu@london.aru.ac.uk