Integrating Generative AI in Academic Assessments: Practical Recommendations on AI Usage and Enhancing Knowledge Sharing

Varun Gupta[1,2,3,*] and Abel Nyamapfene[1]

[1] Centre for Engineering Education, University College London (UCL), London, UK.
[2] Multidisciplinary Research Centre for Innovations in SMEs (MrciS), Gisma University of Applied Sciences, Potsdam, Germany.
[3] Department of Economics and Business Administration, University of Alcala, Alcalá de Henares (Madrid), Spain.

[*]Corresponding author contact: varun.iit13@gmail.com

Abstract

This article presents two strategies for integrating GenAI into assessments: the first strategy involves using the GenAI Integration Rubric to guide ethical AI usage, while the second strategy focuses on the Experimentation Outcome Template to document lessons learned and drive continuous improvement through Monitoring & Evaluation (M&E). These strategies provide practical implications for academic institutions seeking to integrate GenAI into assessments, supporting the reskilling of students while maintaining academic integrity.

Keywords: Generative Artificial Intelligence (Gen AI); academic assessments; AI Contribution Assessment rubrics; Monitoring & Evaluation (M&E); AI experimentations; Experimentation Outcome Template.

Introduction

The integration of Generative AI (GenAI) into academic assessments offers both opportunities and challenges. While GenAI can enhance students learning experience (Wang et al., 2023) and help them focus on higher-order skills like critical thinking (Kizilcec et al., 2024), it also raises ethical concerns, particularly around academic integrity, for instance assignments with AI-generated content submitted as students' own work (Cotton et al., 2023; Dwivedi et al., 2023; Lee et al., 2024; Kizilcec et al., 2024). To harness GenAI's benefits and address its risks, academic institutions must redesign courses and assessments to help students experiment with its capabilities, enhance and apply their critical problem-solving skills, and ensure the integrity of their learning (Gupta and Gupta, 2024). In era of GenAI, the assessments need to be redesigned, focusing on authentic, real-world tasks rather than traditional knowledge recall

(Gruenhagen et al., 2024). In the redesign process, alongside formulating challenging problems for students, assessment briefs should include detailed guidance on the appropriate use and acknowledgment of GenAI in assessments. Gruenhagen et al. (2024) based on a survey of 337 Australian university students, revealed that over a third have used chatbots like ChatGPT for assessment help, often not perceiving it as a violation of academic integrity. Evaluating education programs is vital to measure their impact, ensure lasting changes, and align interventions with the evolving context (Fernández-Díaz et al., 2017). The integration of GenAI must be continuously monitored and evaluated to assess its impact on students by measuring both outputs (e.g., academic performance) and outcomes (e.g., improved ethical reporting and problem-solving skills).

Gupta and Nyamapfene (2024) reported that UCL promotes ethical GenAI use through flexible guidelines in its academic manual. UCL's three-tiered framework for AI use in assessments—prohibition (Category 0), assistive (Category 1), and integral (Category 2)—allows faculties to adapt based on specific needs. Similar to institutions like Harvard, UCL decentralizes decision-making, granting faculty full autonomy to determine whether to allow GenAI and to what extent. Assessments are designed to ensure academic integrity, with platforms like the Generative AI Hub sharing technology information and best practices. To enhance these efforts, faculty can use tools like the GenAI Integration Rubric and Experimentation Outcome Template to align assignments with AI contribution assessments, monitor and evaluate outcomes, and report the evaluation results in standardised way.

These tools are designed based on researchers' experiences with GenAI integration in courses, ongoing research in this domain, as well as the analysis of AI integration in course assessments at institutions like UCL and Gisma University of Applied Sciences. The aim is to provide practical tools that offer tangible value to academic institutions seeking to integrate GenAI ethically into their assessments, ensuring responsible usage while supporting student reskilling and enhancing learning outcomes.

Furthermore, sharing knowledge across the academic community in a standardised format is essential for refining best practices for GenAI integration. Faculty can learn from each other's experiences, adapting strategies to their specific courses and contexts, thereby

enhancing the overall effectiveness of GenAI in assessments. The results of GenAI integration in courses should be evaluated in terms of both the outputs achieved and the outcomes attained. This highlights the importance of integrating Monitoring & Evaluation (M&E) with the experimentation conducted during the integration of these technologies into the courses.

Aligning Assignments with the CRediT Framework for AI Contribution Assessment rubrics (*GenAI Integration Rubric)*

Category 2 (Assistive role) and Category 3 (Integral role) assessments allow the use of GenAI but are designed to ensure the majority of the work remains AI-proof. To help students understand where GenAI can be used, avoided, and acknowledged while ensuring transparency and academic integrity, assessment briefs should be designed accordingly, and instruction guides could be provided. Assessment briefs include instructions defining the scope of GenAI usage. However, it would be valuable for students to receive more detailed guidance from faculty, outlining practical use cases (and potential misuse cases) of GenAI, mapped to the distinct phases of the research process—such as conceptualization, data curation, analysis, and reporting.

The AI contribution assessment rubrics, as outlined in Gupta and Gupta (2024), are adapted into the *GenAI Integration Rubric* (Table 1). This rubric clarifies the 14 roles GenAI can play in research-based assignments, as outlined in the CRediT taxonomy, such as Conceptualization, Data Curation, and Formal Analysis. It defines the roles GenAI can play, outline activities where it can be applied for each role, identify activities that could lead to misconduct, and provide guidelines for acknowledging AI contributions for each role and activity. This ensures students understand the boundaries of AI use, promoting transparency and preventing "unfair advantage," so that the assessment remains the student's own work.

The GenAI integration rubric requires faculty, when designing assessment briefs, to create rubrics that guide students in the following ways: First, by providing specific examples of how GenAI can be used for each role in the CRediT taxonomy, such as using AI for literature reviews or idea generation, ensuring students get insights about possible areas of the assessment where they could try using the technology. Second, by providing specific examples

of activities for each role where AI usage could be considered academic misconduct, such as using AI to generate entire sections of a paper with little to no original contribution from the student. Third, by specifying whether acknowledgment is required for each activity for each role, guiding students on when and how to properly credit their use of AI, for example, in brainstorming or preliminary research. Finally, a rating column allows students to assess the level of AI involvement in their work—Low (*less than 20%)*, Medium *(20 to 50%)*, or High (*above 50%)*—ensuring transparency and adherence to academic integrity. The rubrics will help define the ethical and responsible use of GenAI in higher education and assessments, addressing the challenge highlighted by Gruenhagen et al. (2024) of establishing clear policies for AI integration in academia. The three-tiered categorization of GenAI use in assessments, as suggested in the UCL guidelines, is utilized to illustrate the application of the GenAI Integration Rubric and Experimentation Outcome Template tools. However, these tools are flexible and can be customized to align with the specific categorization frameworks of individual academic institutions. The GenAI Integration Rubric for a sample Category 2 assessment, requiring students to conduct market research for a startup expanding into a foreign country using GenAI, market research tools, and primary research, is given in Table 1.

Table 1: GenAI Integration rubric for a sample Category 2 assessment *(adapted from Gupta and Gupta, 2024, reused with the permission of the Publisher. The figure remains under the standard IEEE license)*.

AI Contribution role	**Questions**	**Example Activities that could benefit from GenAI**	**Example Activities that could result in misconduct *(applicable for category 2)***	**AI Acknowledgement required?**	**Rating** **Low *(less than 20%)*** **Medium *(20 to 50%)*** **High *(Above 50%)***
Conceptualization	*Did you use GenAI to define the research problem or objectives?*	Use GenAI to brainstorm and get directions for the possible research objectives and hypotheses related to market entry.	GenAI generating the entire problem statement without further research/investigation.	Yes	Low, Medium, High
	Did you use GenAI to identify the research	Use GenAI to help identify relevant research questions based on market data and trends.	Using GenAI to generate hypotheses or research questions without any critical	Yes	Low, Medium, High

	questions or hypotheses?		review, further investigation or personalization by the student.		
	Did you use GenAI to develop the initial idea or concept for the project?	Use GenAI to generate an initial framework for the market research project based on initial business goals, which is then further refined through rigorous research.	Relying entirely on GenAI to develop the project idea, with no critical input or further research.	Yes	Low, Medium, High
Data curation	*Did you use GenAI for data management tasks, for instance, organizing, cleaning, describing, enhancing, storing, and preserving data?*	Use GenAI tools to clean and organize secondary data collected from various sources for the market analysis.	Using GenAI to organize data without understanding its context or misusing data for misleading conclusions.	Yes	Low, Medium, High
Formal analysis	*Did you use GenAI to apply statistical or computational techniques to analyze data?*	Use GenAI to analyze market trends, consumer behavior patterns, and competitor data.	Using GenAI to perform all analysis without understanding or interpreting the results or further validation of the analysis.	Yes	Low, Medium, High
	Did you use GenAI to identify the patterns, trends, or relationships in the data?	Use GenAI to highlight trends and correlations within the market data.	Relying solely on GenAI's identification of patterns without questioning or cross-checking.	Yes	Low, Medium, High
	Did you use GenAI to synthesize the data collected from multiple sources into meaningful insights?	Using GenAI to identify common themes across survey responses and industry reports for actionable insights followed by further validation.	Using GenAI to generate analysis without validation and deeper analysis of the insights.	Yes	Low, Medium, High
Investigation	*Did you use GenAI to perform data collection either primary or secondary?*	Use GenAI to gather insights from online databases, academic articles, or industry reports.	Substituting primary data collection (e.g., surveys, interviews) with GenAI-generated data without real-world validation.	Yes	Low, Medium, High

	Did you use GenAI to do a literature survey to inform the investigation?	Using GenAI to summarize key findings from multiple academic papers to identify gaps in the existing research.	Relying solely on GenAI-generated summaries without verifying the original sources, leading to misrepresentation or omission of critical information.	Yes	Low, Medium, High
	Did you use GenAI to conduct experiments, surveys, or observations?	Using GenAI to design survey questions tailored to specific research objectives and analyze initial survey responses to identify trends.	Using GenAI to fabricate survey responses or experimental data instead of collecting genuine primary data. This also includes designing research instruments without grounding them on insights from different sources, for instance literature.	Yes	Low, Medium, High
Methodology	*Did you use GenAI to design the overall approach or methodology for the project?*	Use GenAI to find ideas for the research methodology for market research.	Using GenAI to define the entire research methodology without reasoning about their alignment with the nature of the research questions/objectives.	Yes	Low, Medium, High
	Did you use GenAI to select appropriate research methods and techniques?	Using GenAI to explore and compare various research methods (e.g., qualitative vs. quantitative) and techniques suitable for the market research project.	Relying solely on GenAI to determine the research method without critically evaluating its appropriateness for the research context or objectives.	Yes	Low, Medium, High
	Did you use GenAI to design research study protocols and procedures for data collection and analysis?	Using GenAI to draft initial protocols for surveys, interviews, or data collection procedures, which are then reviewed and refined by the researcher.	Copying AI-generated protocols without validating their relevance, feasibility, or ethical compliance for the specific research context.	Yes	Low, Medium, High
Software	*Did you use GenAI to develop any software application that was used in any*	Use GenAI to design custom software or tools for data analysis (e.g., creating visualization tools).	Relying entirely on AI-generated software without testing or understanding its operation.	Yes	Low, Medium, High

	research activity, for instance, data analysis, modeling, or visualization?				
	Did you use GenAI to implement the algorithms or working software for specific tasks?	Leveraging GenAI to assist in coding or implementing algorithms for analyzing market research data, such as clustering customer segments.	Relying on AI-generated code without testing or validating its functionality, leading to inaccurate or misleading research results.	Yes	Low, Medium, High
	Did you use GenAI to test and debug software components that are the outcome of the research or those required by the research?	Using GenAI tools to identify and debug errors in software components developed for analyzing market data or visualizing research findings.	Allowing GenAI to automatically "fix" bugs without verifying the corrections, which may introduce new errors or compromise the software's intended functionality.	Yes	Low, Medium, High
Validation	*Did you use GenAI to check the accuracy and reliability of research findings, for instance by performing research replications?*	Use GenAI to cross-verify data and findings from multiple sources or simulations.	Relying on AI validation alone without cross-checking with real-world or expert sources.	Yes	Low, Medium, High
	Did you use GenAI to verify the validity of assumptions and methodologies used in the project?	Using GenAI to cross-check assumptions made during market research, ensuring that they align with existing data and trends from other reliable sources.	Relying on GenAI to validate assumptions without critically analyzing the outputs, potentially resulting in unverified or unsupported claims that undermine the integrity of the research.	Yes	Low, Medium, High
Visualization	*Did you use GenAI to create or redesign the visual representations of data or research findings?*	Use GenAI to create graphs, charts, and visuals to represent the research data.	Relying on AI for all visualization without analyzing or adjusting the visual outputs.	Yes	Low, Medium, High
Writing – original draft	*Did you use GenAI to draft full or portions of the*	Use GenAI to help write drafts of specific sections	Using GenAI to generate an entire draft without substantial	Yes	Low, Medium, High

	version of research papers, reports, or other project documents?	of the market research report.	input or edits from the student.		
Writing – review & editing	*Did you use GenAI to review the research documents for clarity, coherence, and accuracy, for instance, improving expressions?*	Using GenAI to review drafts of the research report for clarity and coherence, ensuring that the language is clear, well-structured, and flows logically.	Relying solely on GenAI to rewrite or substantially modify sections of the research document without applying the necessary critical thinking and original analysis, which could lead to excessive reliance on AI-generated content and reduce the originality of the student's work.	Yes	Low, Medium, High
	Did you use GenAI to edit the document for grammar, spelling, and punctuation errors?	Use GenAI to check the document for grammatical, syntactic, and structural improvements.	Using GenAI for the majority of editing work without conducting a personal review of the document that changes substantial portions of student's original work.	Yes	Low, Medium, High
	Did you use GenAI for the feedback and suggestions for improvement to enhance the quality of the writing?	Using GenAI to receive feedback on the clarity, coherence, and structure of the writing, which can suggest ways to improve the flow and organization of the document.	Relying entirely on GenAI for feedback and suggestions without critically evaluating the recommendations or applying them independently, which could lead to over-dependence on AI and undermine academic integrity.	Yes	Low, Medium, High

The use of AI tools can sometimes lead to misconduct, depending on how they are utilized. Table 1 lists some examples per GenAI contribution roles, more examples could be identified and added. For example, using GenAI for the majority of editing work without conducting a personal review of the document, particularly when such edits significantly alter the original work, would constitute misconduct. In contrast, using GenAI for minor edits,

such as correcting grammar, and syntax, or making small structural improvements, may not require acknowledgment—provided the changes are reviewed personally and do not substantially modify the student's original work.

However, whether AI usage requires acknowledgment can vary based on institutional or publisher policies. For instance, when rubrics are designed for student scholarly work, for instance, publications intended for a research degree, they must align with the publisher's policies as well as academic policies, particularly if the work is intended for publication.

To navigate this complex landscape, faculty should provide important misconduct areas in the GenAI Integration rubric, and the students should further identify patterns of acceptable and unacceptable AI use, drawing on the provided examples, other observations, and guidelines. Engaging in discussions with faculty members can help clarify the ethical boundaries and ensure the responsible integration of AI tools in their academic work.

The GenAI Integration Rubric is flexible and can be customized. Faculty can adjust the questions for each AI contribution role, modify example activities for AI usage or misconduct, and update AI acknowledgment guidelines in line with evolving university AI guidance/policies. It is also important to refer to the university's data classification policy (or existing classifications) to identify which data is sensitive and which is not (Gupta, 2024). This helps faculty align the populated GenAI Integration Rubric for the course assignment with these policies.

M&E for Reporting and Sharing Lessons (Experimentation Outcome Template)

Integrating GenAI into course assessments is an ongoing experimental process *(AI Experiments)* with unique challenges and learning opportunities. As the technology evolves, faculty must continuously monitor, evaluate and document not only on outputs, like academic performance, but also on outcomes, such as ethical usage of AI, improvements in problem-solving skills, and students' critical, informed engagement with the technology. These outputs and outcomes, recorded using the Experimentation Outcome Template (Table 2), should be shared with peers through platforms like the Generative AI Hub at UCL, enabling faculty to

exchange lessons, discuss what worked, and share best practices for effective and ethical AI use in education.

Table 2: Experimentation Outcome Template.

Section	Description
Assignment Description	Provide a brief description of the assignment, including the nature of the problem students are required to solve as part of the assessment.
Category of Assessment	Category 2 *(Assistive role)* or 3 *(Integral role)* as per UCL GenAI guidelines.
GenAI Usage	Explain how GenAI is supposed to be used by students, e.g., for data analysis, formal analysis, etc. Link back to the *GenAI Integration Rubric (Table 1).*
Expected Outputs *(Evaluation Indicators)*	Describe what outputs are expected from the successful adoption of GenAI by the students. Examples could include individual grades in the course, GenAI usage reported, etc.
Expected Outcomes *(Evaluation Indicators)*	List expected short-term outcomes, e.g., increased understanding of GenAI tools, enhanced ability to apply GenAI, confidence in ethical usage of the technology, and increased ability to do critical analysis of AI outputs with the insights gained through the application of primary and secondary research, etc.
Challenges/Issues	Identify any challenges or issues encountered during the experiment, e.g., difficulty in maintaining academic integrity, students struggling with ethical AI use, usage of AI humanizing technologies, or any requirement of instructional material, etc.
Recommendations for Peers	Provide insights and suggestions for peers who might want to integrate GenAI into their courses. Focus on best practices, lessons learned, etc. Also mention to what type of courses the recommendations will apply and what courses will find it hard to apply them.
Follow-Up Actions	State what follow-up actions are planned, such as modifying assignments, conducting further experiments, medium- and long-term outcome evaluations, etc.
Survey Information	Mention which indicators are accessed through surveys and which ones through course evaluation results. If possible, provide link to survey questionaries and link to aggregated data *(removing personal data).*

Collaborative sharing fosters peer learning, allowing faculty to adapt insights from each other's experiments to improve their own courses. While lessons from one course may not apply directly to another, they offer valuable guidance for common challenges. M&E requires identifying indicators for outputs and outcomes and conducting surveys where necessary. Some indicators, however, may not require a survey, as the evaluation results themselves will be sufficient to address them.

Challenges may include scalability as student numbers increase and ensuring consistency across courses and departments. Adapting rubrics and templates to diverse course structures may require significant effort, and as the number of templates grows, leveraging collective insights can become more complex. However, the proposed rubrics and templates will provide clear GenAI usage guidelines, foster an experimentation culture, and support M&E-driven reporting of outcomes.

Concluding remarks.

This article presents practical recommendations, such as the GenAI Integration Rubric and Experimentation Outcome Template for the academic institutions aiming to balance innovation with academic integrity and effective learning practices. As GenAI evolves, the skills and lessons learned can be applied to future innovations, enabling faculty to experiment more effectively with emerging technologies and ensure continuous improvement in academic practices.

References

Wang, S., Sun, Z., & Chen, Y. (2023). Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Education and Information Technologies*, 28(5), 4919-4939. https://doi.org/10.1007/s10639-022-11338-4

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., ... & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers*

and Education: Artificial Intelligence, 7, 100269. https://doi.org/10.1016/j.caeai.2024.100269

Cotton, D. R., 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

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper:"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., ... & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. https://doi.org/10.1016/j.caeai.2024.100221

Gruenhagen, J. H., Sinclair, P. M., Carroll, J. A., Baker, P. R., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, 7, 100273. https://doi.org/10.1016/j.caeai.2024.100273

Fernández-Díaz, M. J., Rodríguez-Mantilla, J. M., & Jover-Olmeda, G. (2017). Evaluation of the impact of intervention programmes on education organisations: Application to a Quality Management System. *Evaluation and program planning*, 63, 116-122. https://doi.org/10.1016/j.evalprogplan.2017.04.005

Gupta, V., & Nyamapfene, A. (2025). "Generative AI in Universities: Practices at UCL and Other Institutions, and the Path Forward", *Internet Reference Services Quarterly*, 29(1). https://doi.org/10.1080/10875301.2025.2453461.

Gupta, V., & Gupta, C. (2024). Navigating the landscape of AI-generated text detection: Issues and solutions for upholding academic integrity. *Computer*, 57(11), 118-123. https://doi.org/10.1109/MC.2024.3445068.

Young, A., Chandler, J., Norris, C., & Prevatt-Goldstein, A. (2024). Generative AI and academic skills support at UCL: an institutional approach. *Journal of EAHIL*, 20(2), 16-19. https://doi.org/10.32384/jeahil20620

Gupta, V. (2024). AI experimentation policy for libraries: Balancing innovation and data privacy. *Public Library Quarterly*. https://doi.org/10.1080/01616846.2024.2445356

Acknowledgement

The authors would like to express sincere gratitude to all the participants who generously contributed their time and insights to this research study.

Funding

This research is funded by the Faculty Research and Innovation Grant of Gisma University of Applied Sciences, Potsdam, Germany (Project ID: RGVG_02, Project Acronym: DIGEnt). This publication solely reflects the views of the authors, and the funding agencies cannot be held responsible for any use that may be made of the information contained therein. The funders had no involvement in the execution of the study protocol and reporting, including data collection, analysis, decision to publish, or preparation of the manuscript communicating research findings.

Ethical Approvals

The conducted study was approved by the Institutional Review Board of Gisma University of Applied Sciences, Potsdam, Germany under protocol number 010424. The research project has also obtained clearance under the Academic Technology Approval Scheme (ATAS) of the UK Government's Foreign, Commonwealth, and Development Office (FCDO) dated 01st November 2024.