RESPONSIBILITY, TRANSPARENCY AND TASK DESIGN IN THE AGE OF GENERATIVE AI: A FRAMEWORK FOR ACADEMIC INTEGRITY IN HIGHER EDUCATION

Mariia HALISHEVSKA

* Taras Shevchenko Kremenets Regional Academy of Humanities and Pedagogy, Kremenets, Ukraine
* Corresponding author e-mail: maria.pasevych98@gmail.com

Abstract

The article examined how responsibility, transparency and assessment design can be reconciled when generative artificial intelligence (AI) becomes a routine instrument of higher education teaching and research. The study was conceptual and analytical. It synthesised peer-reviewed literature on academic integrity, AI-text detection and generative AI in education, and compared it with practical experience reported in four presentations delivered by Ukrainian university teachers. Three results were obtained. First, the analysis showed that responsibility for AI-assisted work cannot be delegated, because AI systems have no volition, and therefore integrity remains a property of the human author. Second, transparency and integrity were distinguished as separate dimensions. A four-quadrant model demonstrated that a bare statement "AI was used" does not disclose the role or depth of AI involvement. Third, a model of four delegation zones was described. It allocated tasks to the teacher alone, to manual work, to routine delegation and to a "golden zone" of AI-supported tasks that require professional judgement. The evidence also indicated that probabilistic AI detectors are not a sufficient basis for integrity decisions, and that assessment tasks should rely on live presence, personal experience and visible progress. The article concluded that institutional policies should regulate not only disclosure of AI use but also the role of AI and human accountability for the final product.

Keywords: generative artificial intelligence, academic integrity, attribution, postplagiarism, assessment design, higher education

1. Introduction

Generative artificial intelligence (AI) has moved within a few years from an experimental curiosity to an everyday tool of teachers, researchers and students. Early debates focused on whether such tools should be banned or admitted. Increasingly, the question is how to use them responsibly. Reviews of the educational potential of large language models point to real benefits, such as personalised support and reduced teacher workload. They also stress that these benefits depend on competent human oversight and critical thinking on the part of users (Kasneci et al., 2023). A human-centred approach to generative AI has likewise been the starting point of international policy guidance (Miao and Holmes, 2023).

This shift changes the environment in which academic integrity is realised, although the values of integrity themselves remain stable. Eaton (2023) described the emerging situation as the postplagiarism era, in which advanced technologies become a normal part of teaching, learning and communication, and historical definitions of plagiarism no longer apply in full. Three questions follow. Who is accountable for the result of human–AI interaction? How should the role of AI be disclosed? How should teaching and assessment be organised so that AI supports learning rather than replaces it?

The aim of the article was to propose a coherent framework that answers these questions by combining theoretical literature with practitioners' experience. The objectives were: (1) to clarify the relationship between ethics, integrity and responsibility in human–AI interaction; (2) to distinguish transparency from integrity in the attribution of AI use; (3) to assess the reliability of AI detectors as an instrument of integrity control; and (4) to describe principles of task design and task delegation that keep the human at the centre of education.

2. Materials and methods

The study was conceptual and analytical. Two bodies of material were used. The first consisted of peer-reviewed publications on academic integrity in the AI era, on the performance of AI-text detectors, on AI hallucination and on assessment frameworks. The second consisted of four unpublished practitioner presentations by Ukrainian university teachers and teacher trainers. They concerned academic integrity and attribution of AI use, the practical use of AI in lesson preparation, source work and analysis of student writing, and the selection of AI tools for research and visualisation. The presentations were treated as illustrative evidence of practice and not as a source of generalisable empirical data.
The method comprised thematic analysis of the materials, comparison of practitioners' recommendations with the published literature, and construction of two descriptive models, the transparency–integrity matrix and the delegation zones. Where practitioner claims could not be supported by peer-reviewed evidence, they were presented as practice-based observations.

2.1 Limitations: The study did not include an empirical survey or experiment. The proposed models are therefore analytical constructs that require validation in further research.

3. Results and discussion

3.1 Responsibility and volition: Ethics may be understood as a system of norms that distinguishes right from wrong conduct in a given sphere of activity, while integrity is the conscious and deliberate commitment to follow those norms. The element of volition is decisive. An AI system follows instructions and can be asked, within a prompt, to respect certain constraints, but it has no will of its own and cannot make moral decisions. In every interaction there are therefore two parties: the human who formulates the request, and the system that responds. Only the first can bear academic, ethical and scientific responsibility.
This position is consistent with one of the tenets of postplagiarism, according to which humans may relinquish control over parts of a task but not responsibility for it (Eaton, 2023). The point has a practical basis. Language models can generate plausible but false statements and fabricated references, a phenomenon known as hallucination (Ji et al., 2023). When such errors enter a text, the existence of an algorithmic source does not reduce the user's accountability. The user must verify the output, evaluate it critically and decide whether to use it. For the same reason AI is best treated as an assistant and not as a co-author of research. The researcher formulates the ideas, determines the research logic, interprets the results and draws the conclusions.

3.2 Hybrid writing and the limits of the original/copied distinction: Hybrid writing, in which text is produced through interaction between a person and an AI system, is expected to become normal (Eaton, 2023). Recognising it does not weaken the demand for accountability, but it does change what must be examined. Policies that reduce AI to a yes-or-no question become inadequate, because the degree of AI participation varies widely. Assessment frameworks have responded to this. The AI Assessment Scale, for example, defines five levels of permitted AI involvement, from no use at all to full use, and thereby allows a task to state explicitly what role AI may play (Perkins et al., 2024).
3.3 Attribution: transparency versus integrity: The statement "AI was used in preparing this work" records only the fact of use. It says nothing about whether AI corrected grammar or generated a large part of the text, proposed the structure, or interpreted the results. These situations differ fundamentally, even though the same sentence may describe them all. Attribution should therefore move from declaring use to describing the role of AI. Approaches

discussed in practice include graded labels of AI involvement, descriptions of who was the primary creator of the material and how many interactions took place, and structured declarations of tools, methods and tasks delegated to AI. No universal form exists, and the choice depends on the discipline, the type of work and the rules of the institution or publisher. Transparency, however, is not identical to integrity. A person can disclose AI use in great detail and still hand over an excessive share of the intellectual work. Conversely, a person can work honestly and responsibly and fail to disclose AI use. Table 1 summarises the resulting four combinations.

Table 1. Transparency and integrity of AI use: a four-quadrant model

	Low transparency	**High transparency**
High integrity	Responsible use without disclosure: the work is sound, but readers cannot see the role of AI	Desired state: the role of AI is disclosed, the human retains thinking, verifies the output and bears responsibility
Low integrity	Concealed and uncontrolled delegation: the highest risk	Detailed disclosure of excessive delegation: openness without responsible use

The model shows that policies which regulate only disclosure address one dimension of the problem. A policy may require declarations of authorship, a list of tools and a description of methods, and may even require a log of how the work developed. Without standards for the scope of acceptable delegation and for human accountability, it remains incomplete. A simple three-level logic can supplement it. Some uses are acceptable assistance, some require caution and explicit limits, and some carry a high risk of dishonest use. In every case three questions remain decisive: who thought, who verified, and who is responsible for the final product.

The same logic applies to AI-generated images in scientific and educational publications. Their acceptability depends on the policy of the particular journal or publisher, and the author must disclose how the image was produced. Particularly important is whether it was created from the author's own materials and instructions, or whether the system itself formed a significant part of the visual content.

3.4 AI detectors as an instrument of integrity control: The temptation to rely on detectors is strong, but the evidence does not support it. A systematic test of publicly available detection tools found that they were neither accurate nor reliable enough for use as evidence of misconduct (Weber-Wulff et al., 2023). Detectors have also been shown to misclassify writing by non-native English speakers as AI-generated (Liang et al., 2023). A detector percentage expresses a probability and not an established fact of authorship. It may produce both false positives, for example for formulaic human prose, and false negatives, for text produced or edited with AI.

Consequently, a detector score should not serve as the sole ground for accusing a student or a teacher. At most it is one signal for further analysis, which must also consider the process of creating the work, the sources, the author's contribution and other available evidence. An academic culture should provide procedures for disputed cases and should promote responsible use of AI instead of punishing the mere fact of its use. This is also the more realistic position, since prohibition alone does not remove a technology that is already part of the educational and scientific environment (Cotton et al., 2024).

3.5 Task design instead of policing: The practitioner material converged on one observation: if a task can be completed without thinking, without presence and without personal experience,

it will most probably be completed by AI. The remedy lies in changing the design of tasks. Three supports of an AI-resilient task were identified (Table 2).

Table 2. Supports of an AI-resilient assessment task

Support	**Content**	**Example**
Live presence	Direct interaction in which the student demonstrates understanding	Two-minute oral defence of a written assignment in a seminar
Personal experience	Information not available in public sources	Interview with a practitioner, data from own practice, material discussed in a previous class
Visible progress	Assessment of the process as well as the product	Drafts, prompts, intermediate results, changes between versions

Where a written task is assigned for home completion without any live component, it is reasonable to assume that AI may be used. A more productive response than prohibition is a combined structure of written product, personal experience, documented process and a short oral defence. The same logic applies to test items generated with AI. A question such as "What do you think about these data?" can be answered without reading the source, so items should be reformulated to require engagement with a specific material or with content discussed only in class. This reflects a broader principle: a task must be judged not only by the correctness of its wording but also by the learning action it actually requires from the student.

3.6 Delegation of teaching tasks: Teachers' main constraint is time, and AI is most defensible as a means of releasing it. The practitioners described a model of four zones, based on two considerations: how far AI can actually help, and how much professional judgement the task requires (Table 3).

Table 3. Zones of delegation of teaching tasks to AI

Zone	**Characteristic**	**Examples**
Teacher only	Requires professional, pedagogical and human judgement	Contact with the group, sensing fatigue, changing a lesson in real time, decisions about individual students
Not worth delegating	Faster to do by hand than to describe in a prompt	Small, one-off tasks
Delegate freely	Routine and preparatory operations	Formatting, drafts, initial search, structuring information
Golden zone	AI is very strong but needs professional control	Generating several lesson variants, analysing groups of texts, preparing materials from verified sources

The governing principle of the golden zone is that AI generates options while the teacher analyses, evaluates, selects and adapts. Practitioners reported that preparing a 90-minute class on a new topic for a large cohort could be reduced from about half a day to roughly forty minutes. They also reported that no generated material was used without editing. Several working rules emerged:

- request several variants rather than one, so that options can be compared and combined;
- describe the discipline, audience, level, duration and constraints in the prompt;
- use search tools that display their sources, and check the primary source yourself;
- use tools that work only with uploaded materials when answers must be traceable to those materials;

- use AI to find patterns across a group's writing, such as typical errors, and not to grade individual students;
- remove names and other identifying data before uploading student work.

These rules are the operational form of the general principle of responsibility. Selection of sources according to the aim of the lesson, adaptation of a generated activity to a real group and judgement about assessment remain with the teacher. They are not tasks that can be passed to a system that lacks knowledge of the specific students and the purpose of the course.

4. Conclusions

The analysis supports four conclusions. First, responsibility for the use of generative AI cannot be transferred to the technology. Integrity requires volition, which only a person possesses, so the human author remains accountable for verifying and adopting AI output. Second, transparency and integrity are different properties. Attribution should describe the role and depth of AI participation and not merely state that AI was used, and institutional policies should regulate both disclosure and the scope of acceptable delegation. Third, probabilistic AI detectors do not provide a reliable basis for integrity decisions, and assessment should instead be built on live presence, personal experience and visible progress. Fourth, in teaching, AI is most valuable when it takes over routine work and generates options in the "golden zone", while pedagogical decisions stay with the teacher.

The principal limitation is the conceptual character of the study and the narrow base of practitioner material. Further research should test the transparency–integrity matrix and the delegation zones empirically, for example by analysing institutional AI policies, student and teacher surveys, and experimental comparison of task designs.

Nomenclature

AI Artificial intelligence

References

[1] Cotton D.R.E., Cotton P.A. and Shipway J.R., 2024. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, Vol. 61, No. 2, pp. 228-239. https://doi.org/10.1080/14703297.2023.2190148

[2] Eaton S.E., 2023. Postplagiarism: transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity*, Vol. 19, Article 23. https://doi.org/10.1007/s40979-023-00144-1

[3] Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y.J., Madotto A. and Fung P., 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, Vol. 55, No. 12, Article 248. https://doi.org/10.1145/3571730

[4] Kasneci E., Sessler K., Küchemann S., et al., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, Vol. 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

[5] Liang W., Yuksekgonul M., Mao Y., Wu E. and Zou J., 2023. GPT detectors are biased against non-native English writers. *Patterns*, Vol. 4, No. 7, Article 100779. https://doi.org/10.1016/j.patter.2023.100779

[6] Miao F. and Holmes W., 2023. *Guidance for generative AI in education and research*. UNESCO, Paris.

[7] Perkins M., Furze L., Roe J. and 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*, Vol. 21, No. 6. https://doi.org/10.53761/q3azde36

[8] Weber-Wulff D., Anohina-Naumeca A., Bjelobaba S., et al., 2023. Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, Vol. 19, Article 26. https://doi.org/10.1007/s40979-023-00146-z