

THE ETHICS OF AI-ASSISTED ACADEMIC WRITING. AUTHENTICITY CRITERIA IN THE EVALUATION OF STUDENTS' ASSIGNMENTS

Oana-Antonia ILIE

“Nicolae Bălcescu” Land Forces Academy, Sibiu, Romania

antonia.ilie@yahoo.com

Abstract: *The increasing integration of generative technology into the academic research landscape has sparked disputes regarding the new implications of AI on academic integrity and the evolving role that authenticity plays in this context. This paper explores the issue of authenticity in AI-assisted academic writing, as well as the potential risks to the value of intellectual creation. While the current legislation on artificial intelligence and copyright is still under development, with countries recognizing the need to adapt existing legal frameworks to accommodate the changes brought about by AI advancements, universities are trying to find new ways of fighting the escalating AI-plagiarism phenomenon. Authenticity is a measure of both integrity and value in academic research, envisaged as proof of researchers' academic honesty, originality of their creation, truthfulness and verifiability of the scientific data. The paper proposes the use of authenticity criteria when evaluating students' written academic assignments, criteria that can help educators to distinguish between original creation and AI-generated content. While acknowledging the technical utility of AI systems, the paper argues that human attribution and authenticity are the only qualities that can grant value to academic writing.*

Keywords: AI ethics, authenticity, AI-plagiarism, human agency, human authorship

1. Introduction

The rapid evolution of AI-powered technology has brought forth undeniable benefits for academic learning and research, but also the challenge of reevaluating the concept of academic integrity. The integration of AI-powered technologies into academic learning has contributed to unprecedented developments in teaching, learning, and research, but it has also led to the escalation of the AI-plagiarism phenomenon. Generative AI platforms and writing applications such as ChatGPT, Grammarly or Quillbot, that offer free grammar checking, translation, and paraphrasing facilities support students by enhancing their language skills and the

overall quality of their written texts. Although AI-powered technology also offers the means to detect this new type of plagiarism, in most cases, the reports generated by universities' official plagiarism-detecting systems do not raise suspicions of unethical scientific research. Traditionally, academic integrity is defined by principles such as *honesty*, *trust*, *fairness*, and *responsibility* in the academic context, serving as standards for educational performance. However, in the recent AI-modeled university landscape, academia is confronted with a new obligation: to ensure and preserve authenticity as the main guarantee of the quality of academic writing.

2. Ethics of AI, a Work in Progress

The ethics of AI is a work in progress, with the main questions revolving around aspects such as the *moral agency*, *autonomy* or *identity* of AI systems. Autonomous machines can be considered moral agents in the sense that they can be designed and programmed to perform in accordance with laws and moral rules, to act as if they were human and morally responsible, while “picking carefully whose values or what values they implement” [1]. But at this point, if one artificial system makes moral judgments according to Kantian principles and the other is obeying utilitarian ethics, it is unclear who will be held responsible for the consequences of their decisions.

The current legislation on artificial intelligence and copyright is still in the development phase, with most countries recognizing the need to adapt the existing legal framework to accommodate the changes brought about by the evolution of generative AI technology.

Human agency and *human authorship* are the two fundamental concepts that set the framework for the ethics of AI in education and scientific research. These principles emphasize the importance of human decision and human involvement in guiding the use of AI to achieve the best results in the academic practice. According to the *Ethics Guidelines for Trustworthy AI* (2019), AI must combine three components: *it should be lawful*, complying with all applicable laws and regulations; *it should be ethical*, as to ensure adherence to ethical principles and values, and *it should be robust*, both from a technical and a social perspective. [2] AI systems must be human-centric, created in a spirit of commitment to humanity and the common good, with the goal of improving human welfare and freedom. As such, the ethical imperatives of AI revolve around the following:

Respect for human autonomy (human-centeredness);

Prevention of harm (safety, security);

Fairness (freedom from bias, discrimination and stigmatization);

Explicability (transparency, traceability, auditability) [3].

The *EU Artificial Intelligence Act* (2024) promotes the uptake of a human-centric and trustworthy AI, acknowledging the levels of risks and ensuring the legal measures to provide the protection of health, safety and fundamental rights of the citizens against the potential harmful effects of such systems [4].

The U.S. *AI Bill of Rights* (2022) provides protection against unsafe and ineffective systems, algorithmic discriminations and abusive data practices. *Algorithmic discrimination* refers to the unjustified, biased or unfair treatment of people based on their race, color, ethnicity, when automated systems make decisions based on such criteria as sex, religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. The law also provides that AI systems developers should provide clear notice and explanations of the systems’ functioning and the generated outcomes, and that people have the right to opt out from such automated systems and be provided a human alternative [5]. The *Algorithmic Accountability Act of 2022* supports the AI Bill of Rights by providing requirements of fairness and accountability for AI automated decision systems used by certain businesses, especially when used to make critical decisions. *Critical decisions* are considered the decisions that have a significant impact on a person's life such as their well-being, opportunities, cost or availability of health care, housing, educational pathways, or financial services [6].

China is a pioneer in AI-related policies with various provisions focused on the development of AI to favor national economy and development, alongside concerns about AI safety, security, and ethics. The *New Generation Artificial Intelligence Development Plan* (2017) represents one of the first efforts to set out a top-level goal of global AI leadership by

2030 [7]. The new Chinese regulations such as the *Administrative Provisions on Recommendation Algorithms in Internet-based Information Services* (2021) are extensive sets of rules created to address user rights and the critical issues regarding the societal, human and economical consequences of AI. They also revolve around an ethical code that requires operators to create means for preventing or reducing the spread of illegal or undesirable content and promote ‘positive content’ that reflects Chinese ‘mainstream values’ and cultivates ‘positive energy’ online [8].

The *World Intellectual Property Organization* (WIPO) has explored the main aspects where AI interferes with intellectual property rights and has issued a *Patent Landscape Report* highlighting potential dangers regarding authorship, patent, authenticity, and copyright [9].

International treaties, such as the *Berne Convention for the Protection of Literary and Artistic Works* and the *U.S. Copyright Office*, mention the necessity of a *human author* for works to be protected by copyright. Although most laws currently recognize human agent as creator and beneficiary of copyright, there is a gradual acknowledgment that the law will need to evolve to include consistent legal guidelines for AI-assisted creation. *A Guide on Generative AI and Copyright* issued by the Korean Copyright Commission together with the Ministry of Culture, Sport and Tourism, states that only a natural person can become an author, thus, AI generated works are not eligible for copyright protection. According to Art 2 (1) “a work is a creative production that expresses human thoughts and emotions”[10]. When evaluating creative creation, only the parts that are considered human creative contribution would be considered in order to be granted copyright. Similarly, the *General Understanding of AI and Copyright* (Japan, 2024) indicate that the copyright eligibility of AI assisted creation will be assessed on a case-by-case basis, on

criteria such as the amount of generated content and input prompts, the number of generation attempts, the selections and any other modifications or refinements applied by the user [11].

EU member states share the view that AI generated content cannot be protected by copyright as *only a natural person can be considered an author*. Such works that combine creative human effort with AI creation will be eligible for copyright only if the human input in the creative process is proven significant [12]. The international community is also analyzing whether copyright protection should extend to innovative AI generated material, but in present, there is a consensus regarding the fact that artificial intelligence systems are tools that can only assist in the process of creation, and that they cannot be entitled authors of copyrighted material.

Despite of the fundamental differences between humans and machines and the concerns regarding AI safety and trustfulness, we are gradually becoming aware that there is nothing that AI-powered technology will not be able to do in the future, except maybe for making use of itself. If at this stage of development generative AI is not considered trustworthy enough, as generally agreed, the trust in human agency is at risk to decrease in the following years, if not aligned with AI standards, criteria and evaluation systems.

3. Authenticity Criteria for AI-Assisted Academic Writing

In the AI-modeled Academia research landscape, the most challenging ethical problem encountered is less connected with the aspect of autonomy of artificial agents or the risks associated with the lack of human control but with that of authenticity as value of the generated intellectual creation.

The word *authentic* is a multifaceted concept with chameleonic meanings, ranging from conformity to truth and

reality, to alignment with personal values and legal norms. It can indicate something that is original, not a copy or imitation, an entity or creation that is truthful and genuine, possessing individuality and identity, whose authorship is indisputable. Additionally it can refer to something that is not forged or falsified and, in a legal context, to documents that have been drawn up in compliance with all legal formalities.

While the terms *authentic* and *original* are used interchangeably, and more often function as synonyms, they do have distinct meanings. The concepts associated with the term *original* in dictionaries varies from “initial”, to “innovative”, “unconventional” and “eccentric”. An original work is independent in thought, innovative, and marks a departure from tradition, while authenticity is more often associated with the ideal of conformity to truth and identity. Thus, a work of art can have original traits without being authentic, in case the work is partly falsified, or if it is misattributed. Conversely, a work can be authentic without being original or innovative. While a lack of original contribution does not have legal consequences, a lack of authenticity—such as plagiarism, document falsification, or product counterfeiting—is punishable by law.

Authenticity means living to the laws of one’s own being and choice and represents an ideal sought by individuals and groups as part of the process of becoming [13]. Thus authenticity is related to exemplarity and self-realization: “An identity attains authenticity when it realizes to an outstanding extent the normative core that it sets as its guiding ideal” [14]. Ultimately it is an evaluative concept that establishes the value of intellectual creation as enacting the set of qualities required in order to represent an ideal or exemplary work.

In academic context, *authentic* is synonym with *true*, *original*, *first source*, *natural*,

organic, as opposed to *false*, *fake*, *second source*, *artificial*. Authenticity is a measure of integrity in academic research, indicating the sincerity and academic honesty of the author, the originality of the intellectual creation, and the truthfulness and verifiability of the scientific data. Being authentic in academic research is avoiding the *hoop jumping machinations* that cause losing the vitality and meaningfulness of the research [15], thus authenticity is rooted in sincerity and spontaneous self-expression rather than in conformity to patterns and social forces.

When plagiarism detection systems do not indicate a significant percentage of similarity, but there are stylistic suspicions about the submitted work, it becomes relatively difficult to weigh the authentic contribution by which the work should be evaluated. Therefore, professors find it challenging to decide whether they should recommend the use of AI tools in students' research and creative writing or firmly discourage it. AI removes the *intellectual effort* from *intellectual creation*, an endeavor essential for the selection and processing of information and for the organic development of the skills outlined in the curricula of the respective disciplines. At its worst, over-reliance, or cases where the mind becomes subordinate to the artificial intelligence discursive patterns, leads to the loss of the individual style which is the main condition for authentic and original creation.

Nevertheless, the potential of AI in education is immense: it can become an integral part of the teaching-learning process, supporting teachers and students and complementing traditional methods. Furthermore, it can be an important part of the knowledge journey, as long as it is used as a tool that assists in this process, rather than as an academic shortcut. While AI is redefining education, teachers are obligated to allocate time for educating students on the costs and dangers related

to academic integrity, emphasizing that a paper written by AI, even if flawless, is not a paper deserving of a high grade. Thus, setting authenticity criteria for academic assessments is an imperative in the new context. Teachers should draw attention to the following marks of authenticity that make a paper unique and valuable, and set them as grading criteria:

1. Nuanced Tone and Emotion: The text must contain nuanced thinking, elements that engage the audience's sensitivity, memory or critical thinking, by using a formal, objective tone, without sounding mechanical, or, an empathetic, informal or humorous tone where appropriate.

2. Personal Style: The text should include elements that define the author, and his personality: ideas, opinions, connections with his own personal experiences and cultural background, stories, parentheses, examples, reflections, conclusions.

3. Natural Phrasing: The text should present a varied structure of sentences, shifts in register, nuance, variations in text formatting (emphasis, capitalization), ideas that follow the natural order of the author's logic.

4. Cultural Sensitivity: The text must respect aspects related to cultural inclusion, avoiding biased information and discrimination based on gender, ethnicity, or socio-cultural context, and present a vision that is fair for all social categories.

5. Original Ideas / Creativity:

The text must not only present bibliographical information and data but also include creative elements, original hypothesis, solutions or ideas, and new perspectives that differentiate the paper from other works on the same subject.

Setting authenticity criteria for academic assessments can also act as a motivational force in learning, giving the student an additional incentive to discover their creative self beyond mechanical reproduction.

Determining precisely what is genuine and what is borrowed in a paper may seem

impossible in practice, but nevertheless, the professor can appreciate the creative effort and the students' progress in their expressive abilities within the given context. If students have managed to connect with their true self, with their own phenomenological or emotional experience, reflect on their own ideas and intuitions, and integrate the knowledge they have acquired into their discourse, these aspects can constitute criteria at least as valuable as those that focus on the cumulative aspects of learning.

4. Conclusion

Though understated in present-day Academic Charters, the value of *authenticity* must become the anchor of new academic deontological codes that address the emerging challenges posed by AI-generated research.

The present paper proposes five authenticity criteria for evaluating academic assessments: *nuanced tone and emotion*, *personal style*, *natural phrasing*, *cultural sensitivity* and *original ideas/ creativity*, that can be used to distinguish between authentic, genuine work and AI generated creation.

In the academic community there is a consensus that artificial intelligence systems cannot replace the human author as the main contributor to the academic work. Nevertheless, as legislative frameworks concerning AI and copyright evolve, the predominant academic opinion is gradually shifting away from a hostile attitude regarding the use of AI tools in the creative process, towards a more favorable and permissive view.

Although there are disputes over what defines original writing or what makes a creation authentic in the true sense of the word, many now argue that authenticity can also be achieved in AI-assisted creation through conscious and responsible human involvement, as long as the power of technology is balanced with critical thinking and human creativity.

References

- [1] Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Topdown, bottom-up, and hybrid approaches. *Ethics and Information Technology*, 7(3), 149–155.
- [2] European Commission. Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI, 2019, available at https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf.
- [3] European Commission. Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI, 2019, available at https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf, p. 12.
- [4] European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence. *Official Journal of the European Union*, L 1689, 12 July 2024. Retrieved from <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>.
- [5] White House Office of Science and Technology Policy. (2022). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. Retrieved from: <https://www.govinfo.gov/content/pkg/GOVPUB-PREX23-PURL-gpo193638/pdf/GOVPUB-PREX23-PURL-gpo193638.pdf>.
- [6] U.S. Congress. (2022). Algorithmic Accountability Act of 2022. Retrieved from <https://www.congress.gov/bill/117th-congress/house-bill/6580>.
- [7] State Council of China. (2017). New Generation Artificial Intelligence Development Plan. Retrieved from <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>.
- [8] Cyberspace Administration of China. (2021). Administrative Provisions on Recommendation Algorithms in Internet-based Information Services Retrieved from: <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/>.
- [9] World Intellectual Property Organization. Generative Artificial Intelligence: Patent Landscape Report. Geneva: WIPO, 2024 Retrieved from: <https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/index.html>.
- [10] Korean Copyright Commission. (2024). A Guide on Generative AI and Copyright, p.40 Retrieved from <https://www.korea.net/Government/Briefing-Room/Press-Releases/view?articleId=391&insttCode=A260123&type=N>.
- [11] Japan Copyright Office. (2024). General Understanding on AI and Copyright in Japan, p.17 Retrieved from https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf.
- [12] Council of the European Union. (2024). Policy questionnaire on the relationship between generative Artificial Intelligence and copyright and related rights – Revised Presidency summary of the Member States contributions, p.3. Retrieved from <https://data.consilium.europa.eu/doc/document/ST-16710-2024-REV-1/en/pdf>.
- [13] Williams, J. P., & Vannini, P. (2009). Authenticity in Culture, Self, and Society. Routledge, p.3. <https://doi.org/10.4324/9781315261973>.
- [14] Ferrara, A. (2008). The Force of the Example. Explorations in the Paradigm of Judgment. New York: Columbia University Press, p.3
- [15] Jacobs, D. T. (2008). The Authentic Dissertation: Alternative Ways of Knowing, Research and Representation. Routledge. <https://doi.org/10.4324/9780203870501>.