

ORIGINAL ARTICLE

Open Access



Faculty members' use of artificial intelligence to grade student papers: a case of implications

Rahul Kumar*

*Correspondence:
rkumar@brocku.ca

Faculty of Education, Brock
University, St. Catharines, Canada

Abstract

This paper presents the case of an adjunct university professor to illustrate the dilemma of using artificial intelligence (AI) technology to grade student papers. The hypothetical case discusses the benefits of using a commercial AI service to grade student papers—including discretion, convenience, pedagogical merits of consistent feedback for students, and advances made in the field that yield high-quality work—all of which are achieved quickly. Arguments against using AI to grade student papers involve cost, privacy, legality, and ethics. The paper discusses career implications for faculty members in both situations and concludes with implications for researchers within the discourse on academic integrity.

Keywords: Artificial intelligence, Pedagogy, Marking, Academic integrity, Professorial conundrum

Introduction

Robert Coles (1989) stated that case studies are “reservoirs of wisdom” (p. xii)—pools of knowledge that reflect concepts and ideas and bring into focus issues that otherwise would remain concealed. Case analysis can focus our collective attention on corrective actions that can improve future directions and outcomes and also pre-emptively reveal what might happen (or be happening) if such issues are not reported. Moreover, hypothetical cases can spare real personalities from possibly embarrassing or ethically charged situations. Altogether, as Hoy and Tarter (1995) summed up, “case studies have a long tradition of bringing provocative problems from the field to analysis and solutions” (p. xi), which this paper sets out to do.

Calls to use artificial intelligence (AI) in post-secondary education (PSE) are being heralded as opportunities (Martiniello et al. 2020) and equally vociferously are the calls to be cautious (Barnett, 2023; Wingard 2023). In this instance, a plausible hypothetical case can shed light on educational integrity issues and ethical concerns. The case designed and presented here relates to a faculty member rather than a student. This undertaking illustrates the need for policymakers, administrators, and the academy as a whole to focus on all stakeholders to preserve educational integrity, rather than narrowly focusing on students' actions and reforming their attitudes.



Case description

Dr. A.I. Case is an adjunct faculty member who works in a faculty of education affiliated with two publicly funded universities. Dr. Case aspires to be tenured eventually, which will require extensive research and publishing contributions in addition to university teaching experience at both the undergraduate and graduate levels.

Dr. Case's workload includes teaching twice as many courses compared to his tenured colleagues at either of the two universities. Furthermore, teaching across two universities involves considerable time commuting (by car) between home and the two campuses. Again, if Case aspires to secure a tenured position, he must devote considerable time to research and publishing, yet the realities of contingent low-paid work requiring extensive travel and a heavy teaching workload barely afford him a living. Although he contemplates stepping down from his assignments at one of the universities to dedicate more time to research and publishing, his life circumstances do not allow it.

Case has a young family, with its attendant responsibilities. His partner has a shift-work position at a retail store. Their 2-year-old daughter goes to a local daycare centre when her grandparents and parents cannot care for her; it is expensive, but the Case family has no other options. Because of the current fiscal demands on universities, class enrollments in the courses Case teaches are about twice as high as when he attended university. Higher enrollment has meant more marking, which takes up much of Case's time when he is at home. This has caused tension between Case and his spouse, who accuses him of neglecting his familial commitments.

Late one night, during a break in a long marking session, Case posted on social media that he was overwhelmed with his workload, especially marking. Shortly after, he saw a post on his feed from a company that employs artificial intelligence (AI) technology to grade student papers, offering to discretely *help* with his workload. Case was unsure if that was legal or proper or even ethical. He ignored the post and resumed his marking, finishing well past midnight.

Early the next morning, Case decided to explore exactly what the posted AI service would do and what it would entail. He filled out a detailed account creation form that requested: (a) the level of student papers (undergraduate or graduate), (b) the number of papers for each assignment, (c) the number of sets of assignments Case intends to have marked by the AI bot, and (d) yearly commitments. In deciding whether to use this service or not, Case drafted the following table that listed the benefits and pitfalls of using the AI service to grade student papers (Table 1):

Table 1 Categories of possible benefits and pitfalls of using AI to grade student papers

Benefits of using the AI grading service	Pitfalls of using the AI grading service
B1. Discretion	P1. Cost
B2. Time saving	P2. Privacy concerns
B3. Convenience	P3. Legal concerns
B4. Consistent student feedback	P4. Ethical concerns
B5. Adequate quality	P5. Inadequate quality
CB6. Career implications	CP6. Career implications

Potential benefits of the AI grading service

B1. Discretion

In its communication, the AI service provider assured Dr. Case that any personal information collected about Case himself, the details of student papers that he wants to grade, and student records would not be shared with anyone. Dr. Case would not have to interact directly with any company representative, which provided a cloak of privacy that appealed to him. Furthermore, Case's employers did not have to approve nor authorize his use of the service, which reassured him that the service was discrete.

B2. Time saving

The most significant reason Case considered using an AI bot to grade student papers was to save time and restore a healthier work-life balance – an argument that non-human aids can save time have been repeated since Pressey (1926) to more recently by Holmes et al. (2019). Case anticipated that it would help him keep up with his share of family responsibilities, which his spouse accused him of neglecting because of his work schedule and long extracurricular marking sessions at home. Case also reasoned that using the AI grading service would allow him to provide his students with prompter feedback, thereby improving their learning. He also would be able to scaffold their assignments (Bliss et al. 1996). In other words, there was pedagogical merit in using such a service.

Additionally, unlike a conventional enterprise with a brick-and-mortar storefront and traditional weekday business hours, Case's interaction with the AI service provider involved electronic communication with an anonymous agent (i.e., virtual assistant) at any time of the day. Such communication sped up the transaction of setup and subsequent communication between Case and the AI service provider considerably and could possibly make the entire process fast.

From what Case understood, the AI grading tool's user interface would allow different configurations for different assignments for each of his courses. In short, the entire grading process could be automated and "hands-free" once it was set up and configured according to Case's specifications. That alone would save him much (in)valuable time.

B3. Convenience

The entire system was convenient too, offering Dr. Case three options: First, he could upload student papers from within the comfort of his home; all he needed was a computer and a stable internet connection. The second option was even more convenient: Students could upload their assignments directly to the AI server. Third, if Case wanted to use the service on a regular, ongoing basis, he could link the university's learning management system (LMS) for accepting student submissions with the AI provider's repository, but this option would require university involvement. In the latter case, students continue to do whatever they normally do to upload their assignments in their respective institution's LMS, thereby not requiring alterations to the workflow of assignment submission. Because Dr. Case was still unsure of the quality of service and reaction from students, his colleagues, and administrators, he wanted to keep the entire exploration secret and private. After all, there were power dynamics to be considered and delicately navigated. Consequently, Case contemplated the most discrete option (uploading the assignments himself), which still would be quite easy and convenient.

B4. Consistent student feedback

Dr. Case is aware of the value and effectiveness of feedback in improving learning for his students (Al-Bashir et al. 2016). Lately, Case had been worried that the extent of feedback he provided to students varied on many personal factors, such as fatigue, time of day he marked assignments, and turnaround time expectations (Mumford & Atay, 2021). That is, the quality and depth seemed inversely proportional to how tired he was and how much time he could devote to each assignment. Although Case acknowledged that he may have exceedingly high (self-imposed) standards for the provision of feedback, he also reasoned that using an AI bot would provide consistent feedback to students (Ellis 2022), as a result making it fairer than his current practice that was affected by his human limitation (i.e., the need for adequate sleep).

B5. Adequate quality

In addition to consistency, Case wondered about the quality of the evaluations and student feedback. Any other benefits of the AI grading service would be irrelevant if the quality was lacking. In other words, the quality of feedback had to be good. While an interpretation of “quality” may vary based on disciplinary affiliation (e.g. computer science vs. education), type of assignment (e.g., objective answers vs. long essay), instructors’ expectations of student output (e.g., identifying gaps in vs. confirming and building upon students’ knowledge), and the level of work evaluated (e.g. undergraduate vs. graduate work), a principal aspect of assessing quality would be to distill the student work to its main points. Dr. Case is aware that the currently popular AI-chatbot ChatGPT has the ability to address this aspect of quality quite well; it can compare the seminal ideas of the submitted assignments against a set of comprehensive answer keys (ChatGPT 2023a, 2023b) and even tailor a draft response to students. Would the advertised AI grading service automate the human interventions with ChatGPT? Would the quality of descriptive, explanatory, or essay-style answers be similar to or better than those which ChatGPT produces?

CB6. Career implications

Dr. Case believed that using an AI grading service to keep up with marking in large classes might facilitate his goal of eventually securing tenure or at least not hamper his chances. One of Case’s colleagues who had expressed a formal grievance to his employer about the growing workload due primarily to assessment expectations, did not have his contract renewed. Case interpreted (and internalized) that incident to mean he was not to speak up about his untenable workload. He also inferred that by using the AI-service, he would be able to meet his job expectations, which would only help his case of being an agreeable employee and not jeopardize his career aspirations.

Potential pitfalls of the AI grading service

P1. Cost

After filling out the AI grading service’s requested information form, Dr. Case received a price quote, which was not insignificant. If he subscribed to the service, he would have to allocate a substantial portion of his income toward grading student papers. Moreover, the terms of using the service required a valid credit card, which would be charged a

base fee (either monthly, per term, or annually) in addition to per-use cost. Even if Case could offset the cost via an institutional account with his employer, such funds would still fall far short of the cost of using the AI marking service. Moreover, the prices would likely increase over time, and it was unclear if he would be able to terminate the contract without a considerable cancellation fee.

P2. Privacy concerns

Two problematic privacy issues emerged in Case's contemplation of using the AI grading service: his own and his students. Case was concerned about how the information collected about him would be used: Would the company list his name on its website? Would his employers be notified that he was using the company's service? Would students be told that he uses AI technology to grade their papers? What happens to his information if he terminates the contract with the AI company? Would his use of the AI grading service be disclosed to other universities where Case might apply in the future to get a permanent (tenured) position?

In addition, student privacy concerns also loom: How would the company use the submitted student papers? Would the assignments be stored in perpetuity or purged after they were graded? Who would have access to them? What private information would the company collect if the students submitted assignments through the company's portal? If the university's LMS was calibrated so that the AI system could access students' submitted assignments and the graded marks could be transferred to the LMS's grade book without human intervention, such a configuration would ensure the assignment submission process for students would be unaltered. The assignment evaluation process would be automated and convenient for students and faculty alike; however, would such a configuration divulge unintended information about students (e.g., their aggregate grades, full names, and user codes)? Similarly vexing questions emerged when text-matching software solutions, like Turnitin, were widely adopted in the late 2000s (Brinkman 2013; Vanacker 2011). Are these matters of concern in this new context of AI?

P3. Legal concerns

Dr. Case ponders relevant legal concerns about using the AI grading service. If Case used the service without requesting permission to do so from his department or university, would that constitute a contractual or legal violation? Is it assumed that the hired instructor (adjunct professor) would evaluate student assignments without outside help, or is that an explicit contractual obligation? Would only the latter constitute a legal violation? Would AI-system usage to grade student papers be considered within or outside the stipulated limit? At some universities, an expectation is often conveyed informally that the instructor on record will mark a predetermined percentage of students' work. These issues emerge because there is an implied assumption that the course instructor will do the grading (either all or in part).

P4. Ethical concerns

Whether legal considerations are relevant might be debatable, but ethical considerations are pertinent here. If not legally, there is an assumption that those that teach can best assess learning. The root word for assessment is *assidere*, which means to sit

beside (Swaffield 2011). At least two fundamental principles captured in two simple questions need to be deliberated: First, would Dr. Case's use of the AI grading service be the right thing to do? Second, would his use of the AI service be a good thing to do? This distinction relies on Ross's (1930) discussion of the right and the good that underpins two prominent theories of ethics.

The first question concerning the *right* thing to do evolves into doing things based on a principle: In this case, is using an AI service to grade student papers a correct or proper thing to do? Dr. Case reasons that because his employers have not explicitly prohibited the use of such technology, his potential use of the AI grading service would be okay. However, the idea of a *right* action goes beyond the technical absence (or presence) of such guidelines from employers. It is akin to the Kantian deontological idea—the act is only justified insofar as it instantiates a principle—that Case would want everyone to do under all circumstances. That is, would we want all instructors to use AI services to evaluate student papers? The Kantian answer, which advocates principle-based actions, would be that unless we do not object to all instructors using such a service, we should not use it ourselves. That is, the principle has to be universal. Whether such a consensus exists is debatable.

The second question concerning the *good* thing to do construes the evaluation of actions that produce desirable results: In this case, is the use of AI to grade student papers a good thing? The act of using an AI service to grade student papers entails at least two properties: the quality of the evaluation and the other positive features. Assuming that the quality of the evaluations is sufficiently high, which Dr. Case must take at advertised face value since he has not used the service yet, the other positive attributes play a significant part. The papers are evaluated quickly (much faster than what Case can do), and the feedback is received by students sooner, allowing them to learn from and improve upon their previous work in subsequent assignments (Al-Bashir et al. 2016). Pedagogically, it will also allow Case to scaffold assignments such that students receive quick feedback on their work, which can build towards the final capstone project/paper at the end.

Dr. Case's decision to use AI to mark student papers would ensure that the assignments in his courses could build up toward a substantive assignment when timely feedback is provided at each step. In this way, short assignments leading up to a capstone project fulfill Novak's (2009) idea that "the process is the product in the making" (p. 17). That is, the artificiality of essay assignments in courses as a way to produce papers is replaced by organic learning of how to integrate course content to tackle a salient issue and learn how to analyze and communicate it. The end product of this process is still a paper, but it is an outcome of an authentic engagement with the learning process for the student. If using an AI service helps achieve this goal, is it unethical?

In sum, reliance on theories of the *good* suggests there is nothing wrong with using AI to grade student papers if it can be carried out in a timely fashion, with high-quality evaluation and feedback, with convenience, and at a reasonable cost. Reliance on theories of the *right*, in turn, is predicated on whether the use of an AI service is right or wrong. Public sentiment on the matter is unknown; there can be some speculation,

but the current distribution of each position (for or against) has not been identified. Such data are needed to ensure that the principles adopted in policies represent the prevalent informed views.

P6. Inadequate quality

Case's concern about the quality of the work rendered by AI tools was consuming. What if the quality of feedback was inferior, biased, and inaccurate? This was possible because the underlying Large Language Models that form the basis of AI bots like ChatGPT were trained using large datasets from the Internet, which are societal reflections of flaws and have built-in biases and prejudices (Sharma 2023). Would AI-generated comments in the form of feedback run the risk of perpetuating the biases education attempts to remove? Would it tarnish Dr. Case's reputation in the process? In other words, the quality of the feedback across various assignments was still an ongoing concern. Each assignment would act as a prompt, eliciting a response from AI, and unless Case reviewed each response to ensure quality feedback, he would not know. In that situation, using AI might not be time-saving or worthwhile. After all, diminished feedback quality (prone to prejudices and perhaps even offensive remarks) was grounds for the dismissal of Dr. Case from his job and AI from marking student papers.

CP6. Career implications

Three components can contribute to negative career implications for Case. First, is the use of AI technology to mark papers legally permissible? Second, what attitudes are prevalent amongst institutional administrators and colleagues who would decide whether or not to renew Case's contract? Third, if Dr. Case's case gains publicity, the public's positive or negative views on university professors' use of AI to grade student papers would have a tremendous influence on university administrators' decision to renew Case's contract. Dr. Case could potentially lose his current position, never mind advancing his tenured employment ambitions. And given the current misunderstandings among the public about the strengths, limitations, and scope of the current technologies, his prospects seem grim. Beyond Dr. Case's case, public sentiment could also shape the policies and practices adopted in universities. Although when it comes to policies, industry demand would also factor in prominently.

Final thoughts

After having considered the issues, Dr. A.I. Case wondered how he would proceed. Arguments in favour of using AI to grade papers are just as compelling as those against it. Case reasoned that not using the AI system would be a safe bet. But what if he were to use ChatGPT manually to generate student feedback? Would that differ from Case's predicament of whether or not using the AI grading service is more convenient? Is it the transfer of money that makes it problematic? If Case told his students that he would be using AI technology to grade their papers, would it make his action more palatable (or less objectionable)?

As Dr. Case contemplates his dilemma, what additional issues must he consider? Should the decision to use the AI grading system be made by Case alone, or do

the two universities need to provide direction and boundaries? Consider, too, that if Case secured a full-time position at either institution that significantly reduced the time commitments of his dual-institution situation, he might not even consider using the AI system just yet. Perhaps he would confront such questions in the future when all faculty members (tenured or not) have reached some consensus on the topic. A further question that complicates the discussion is the assumption that undergirds the legal and ethical issue: Whose job is it to mark student papers? Is it merely an unarticulated assumption or expectation? Or is it actually a matter of educational integrity that is laid bare because of the disruptive force of artificial intelligence? How this question is settled will be answered in time, and likely the responses will shift over time.

Conclusion

With the release of Dall-E, ChatGPT, and other AI-based services, the discourse on how these technologies will be used in the academy has burgeoned. The geopolitical and economic competition and uncertainty has the public worried and also excited if media (and social media) reports are to be believed (Rainie et al. 2021). Educators will have to place themselves on the continuum between the two poles of: (a) Ignoring AI tools and (b) Educating about AI tools. And on the other perpendicularly intersecting axis (a) Banning AI tools and (b) Whole-heartedly adopting them. The aspiration to ban such technologies is a losing proposition; embracing them and using them judiciously is a better and more productive choice advocate scholars (Eaton et al. 2021). However these prescriptions arrive from student work. Do these dictums still apply to faculty work? The hypothetical case in this paper compels us to examine alternative possibilities of using AI to grade student papers. Embracing such technology means that people like Dr. A.I. Case could use it to mark student papers. Would that trouble our existing notions of educational integrity? Why or why not? Dr. Case can anticipate the delicious experience of mystery and discovery that lay ahead.

Additional questions

The scope of this paper does not permit exploring other questions that emerge from this case. Still, they are worth raising because they can propel further in-depth conceptual and empirical investigations into education integrity. They can also be discussed with students in class. Five salient among them are:

- What are the differences, if any, between applying educational integrity to student matters and faculty matters?
- What is the current professorial mood for using artificial intelligence technology to grade student papers? How does professorial mood differ from the general public and student sentiments? Are there disciplinary variations?
- What would be the implications of using artificial intelligence to mark student papers have on the quality of education, if any?
- How do public sentiments about using artificial intelligence to grade student papers change over time?

- While there are embedded power issues between faculty members and students, there are also power differentials between various levels of professors in the academy that complicate matters. How do power dynamics affect this case? What about the power differential between the professoriate and the administrative cadre? How do power issues complicate educational integrity matters?

Abbreviations

AI	Artificial Intelligence
LMS	Learning Management System
PSE	Post Secondary Education

Acknowledgements

The author would like to thank the Gibson library services for funding the APC. The author would also like to thank the reviewers and the editor-in-chief of IJEL for providing valuable comments on the earlier draft of the article.

Author's contributions

Hundred percent of the work is done by the listed author. The author read and approved the final manuscript.

Funding

The author would like to thank his university library for funding the article publishing cost (APC).

Availability of data and materials

Since this is a hypothetical case, the data and material are not applicable.

Declarations

Competing interests

The author declares no competing interest or conflict of interest.

Received: 31 January 2023 Accepted: 27 March 2023

Published online: 15 May 2023

References

- Al-Bashir MM, Kabir RM, Rahman I (2016) The value and effectiveness of feedback in improving students' learning and professionalizing teaching in higher education. *J Educ Pract* 7(16):38–41 <https://iiste.org/Journals/index.php/JEP/article/view/31077/31909>
- Barnett, S. (2023). ChatGPT is making universities rethink plagiarism: students and professors can't decide whether the AI chatbot is a research tool—or a cheating engine. *Wired*. <https://www.wired.com/story/chatgpt-college-university-plagiarism>
- Bliss J, Askew M, Macrae S (1996) Effective teaching and learning: scaffolding revisited. *Oxf Rev Educ* 22(1):37–61. <https://doi.org/10.1080/0305498960220103>
- Brinkman B (2013) An analysis of student privacy rights in the use of plagiarism detection systems. *Sci Eng Ethics* 19:1255–1266. <https://doi.org/10.1007/s11948-012-9370-y>
- ChatGPT (2023a). "Can you compare pieces of writing?" [prompt]. OpenAI. <https://chat.openai.com>
- ChatGPT (2023b). "Can you extract essential points from a piece of writing?" [prompt]. OpenAI. <https://chat.openai.com>
- Coles R (1989) *The call of stories: teaching and the moral imagination*. Houghton, Mifflin and Company
- Eaton SE, Mindzak M, Morrison R (2021) The impact of text-generating technologies on academic integrity. In: AI & AI. Canadian Association for the Study of educational administration (CASEA). University of Alberta <http://hdl.handle.net/1880/113569>
- Ellis E (2022) The potential of artificial intelligence in assessment feedback. *The Times Higher Education* <https://www.timeshighereducation.com/campus/potential-artificial-intelligence-assessment-feedback>
- Holmes W, Bialik M, Fadel C (2019) *Artificial intelligence in education: promises and implications for teaching and learning*. Centre for Curriculum Redesign
- Hoy WK, Tarter CJ (1995) *Administrators solving the problems of practice: decision-making concepts, cases, and consequences*. Allyn & Bacon
- Martiniello N, Asuncion J, Fichten C, Jorgensen M, Havel A, Harrison M, Legault A, Lussier A, Vo C (2020) Artificial intelligence for students in postsecondary education: a world of opportunity. *AI Matters* 6(3):17–29. <https://doi.org/10.1145/3446243.3446250>
- Mumford S, Atay D (2021) Teachers' perspectives on the causes of rater discrepancy in an English for academic purposes context. *Assess Writ* 48(2021):100527. <https://doi.org/10.1016/j.asw.2021.100527>
- Novak JM (2009) *Invitational leadership*. In: Davies B (ed) *The essentials of school leadership*, 2nd edn. SAGE, pp 53–73. <https://doi.org/10.4135/9781446288290>

- Pressey SL (1926) A simple apparatus which gives tests and scores—and teaches. *Sch Soc* 23(586):373–376
- Rainie L, Anderson J, Vogels EA (2021) Experts doubt ethical AI design will be broadly adopted as the norm within the next decade. PEW Research Center <https://www.pewresearch.org/internet/2021/06/16/1-worries-about-developments-in-ai/>
- Ross WD (1930) *The right and the good*. Clarendon Press
- Sharma S (2023) ETtech explainer: Warnings aplenty against AI chatbot biases, ignore them at your own peril. *The Economic Times* <https://economictimes.indiatimes.com/tech/technology/warnings-aplenty-against-ai-chatbot-biases-ignore-them-at-your-own-peril/articleshow/97842434.cms>
- Swaffield S (2011) Getting to the heart of authentic assessment for learning. *Assess Educ* 18(4):433–449. <https://doi.org/10.1080/0969594X.2011.582838>
- Vanacker B (2011) Returning students' right to access, choice and notice: a proposed code of ethics for instructors using Turnitin. *Ethics Inf Technol* 13(4):327–338. <https://doi.org/10.1007/s10676-011-9277-3>
- Wingard J (2023) ChatGPT: a threat to higher education? *Forbes* <https://www.forbes.com/sites/jasonwingard/2023/01/10/chatgpt-a-threat-to-higher-education/>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

