

Artificial Reviewers: Teaching Academic Writing with ChatGPT

Abstract: This short work is a contribution to the literature on practices involving artificial intelligence in the academic writing classroom. It presents a centerpiece exercise developed for philosophy and cognitive science students that introduces them to the publishing process. It makes use of artificial intelligence (ChatGPT) to generate reviews for short papers written by students, which can then be used as material for students to address – requiring them to make revisions and to reply to the review. It serves the dual purpose of engaging students with their own writing while exercising competences relating to the peer review process. The exercise itself is compelling, and importantly it is easy to execute. No special skills or resources are required on the part of the instructor, making it accessible irrespective of technical ability or institutional resources.

Keywords: Artificial Intelligence, ChatGPT, LLM, Writing, Language, E-learning

While Large Language Models (LLMs) pose many problems from the perspective of written assignments and teaching across higher education classrooms, there is an interesting application of LLMs that instructors can leverage for their own teaching purposes. What is outlined here is a novel use of LLMs to support a written assignment that can be incorporated into academic writing courses as part of units on publishing and peer review. The assignment introduces a faux peer review process, in which students are tasked with responding to a review of their own earlier work. As part of that process, Chat-GPT can be leveraged as a synthetic reviewer, providing students with individualized critical feedback on their work, modeling peer review, and inviting revision from the students. The LLM makes the task of producing individualized reviews for dozens of student essays feasible, which provides students with a more engaging assignment and a more in-depth experience with the revision process.

The contribution of this work is an outline of how this sort of exercise can be structured, how it fits into a larger course design, as well as presenting an example workflow together with a set of anonymized examples of the output. The process itself is simple and easy to replicate. The tools used are readily accessible, with no barriers to entry, financial or otherwise. The result is a more engaging and memorable assignment for students, one that would not be feasible without the support of this kind of LLM.

The first section elaborates on the context of the course together with some of the challenges relating to this type of assignment. It also describes the relevant particulars of the course, including the preceding unit and how it naturally sets up the assignment. The second section notes the required tools and provides the steps followed to generate these artificial reviews along with two examples generated for student papers submitted in the past. The third section presents some basic data concerning the student reception of the exercise, which is followed by a brief conclusion.

1. Course context and background

This novel application of an LLM as an “artificial reviewer” fits into an introductory course to academic writing at the university level. It has seen implementation and refinement in a course designed for undergraduates in philosophy and cognitive science programs organized by the philosophy departments at two universities – for first-year philosophy students and third-year cognitive science students, to be precise. Class sizes in this context have ranged from fifteen to thirty students and one to three classes per semester. This exercise was introduced four semesters ago, in the 2022/2023 academic year.

The course itself covers writing best practices, argumentation forms, the publication process, and publishing ethics (see, e.g., Martinich, 2016; Bailey, 2021). The idea of the course is, naturally, to introduce the students to the basic elements of writing and composition, moving from things like outlines and structure through to style. This is then followed by an extended unit examining different types of tactics and strategies in argumentation and how they can be employed in academic writing. Finally, these basics concerning writing are coupled with the outlet of academic writing – peer review and publication.

The teaching plan is designed with a progression in mind, where the closing exercise for the unit on tactics and strategies in argumentation is a very short written assignment. The students are required to produce a one-page paper where they apply at least two major tactics (e.g., a variety of dilemma and a *reductio*). They are encouraged to choose their own topic and to be creative, with the themes ranging from serious and erudite arguments involving ethics and ontology to playful arguments that intentionally engage in absurd discussions of definitions or tastes. The course then transitions to a unit on peer review and the publication process.

In exploring this process, the course goes through many of the basic concepts surrounding publications, covering things like how journals are structured, who reviewers are, the expected time-frame for publication, etc. This culminates in the presentation of how reviews are to be handled – that is, the best practices when responding to referees. This ranges from simple things like being polite, to explaining how responses should be structured, how changes should be introduced, and when and how to decline to make certain changes.

The intention that I had a long time ago when designing the course was for the short essays the students prepared in the preceding unit on tactics to be used as material for a review that could then serve as an exercise in the present unit on peer review and publication. This would

bring the work of the student in the previous unit back into the fold, confronting them with some potential issues and encouraging them to make changes. It would reinforce the lessons of the preceding unit while also providing them with some experience of the processes at issue in the current unit.

The intention initially was that I could simply do this myself, such that I would prepare tailored reviews for them – which was, in retrospect, unrealistic. There is simply not enough time for an instructor to prepare reviews for so many texts, each of which would need to be carefully structured so as to provide an adequate example of a review while also meaningfully engaging them in some form of revision. The idea that followed was to pair students with one another to quite literally recreate the peer review process – such that one student would write a review for another, and they could then respond to each other. While that is far more economical, it is much more difficult to police this sort of exercise or to ensure good reviews are actually being presented. Moreover, if one student fails to do the assignment, their partner is left with no material at all. These are seemingly minor but ultimately rather serious issues, creating more problems than they solve. For some time, the solution was to simply provide all of the students with one example paper and an example review. They would then all address the same paper and review in their own words, making changes and responding as they saw fit, in line with the course.

However, the solution seems to have finally come in the form of the readily available and accessible tool that is ChatGPT. It is capable of providing a tailored review out of the box, with a basic summary and critical remarks. It does a wonderful job of producing short and professional responses that the students can then be presented with as a review tailored to their specific work. It effectively realizes the initial intention of this particular course progression, as it prompts the kind of reflection noted earlier while requiring the student to produce a working revision and reply. Importantly, ChatGPT does this expediently and at practically no cost to the instructor while also requiring almost no technical skill. This very simple application of AI is an example not just of how AI may make a difference in the classroom, but of how easy it has become to design applications like this – allowing for a tailored experience that better engages students at scale.

Part of what makes this interesting is how it fits into the considerable literature on the use of AI in writing assessment across writing courses, in particular automated writing evaluation tools (see, e.g., Dale & Viethen, 2021; Godwin-Jones, 2022; Woodworth & Barkaoui, 2020; Seufert et al., 2021). The majority of that literature has focused on how evaluation tools can be implemented to improve student writing; however, these evaluation tools are limited in terms of their feedback. Though “intelligent,” they cannot provide the kind of content-oriented review that a more universal LLM is capable of generating, which is the differentiating factor in this particular application. It also distinguishes itself from the ongoing discussion of the significance of these models from the perspective of student use (see, e.g., Chan & Hu, 2023; Mouser, 2024), contributing rather to the relatively small but growing literature of concrete

applications these tools provide us as instructors (see, e.g., Ilieva et al., 2023; Matzakos et al., 2023).¹

While this may all seem modest, the matter of producing exercises like this is no small feat for an academic writing course. This sort of exercise forms a centerpiece assignment that serves to reinforce the competences covered in the unit, to test student competences to that point, and to present a more memorable and compelling experience for the students as they engage with their own earlier work in an extended faux peer review process.

2. Exercise design and procedure

The task is accordingly to take some piece of critical writing produced by the students and to generate reviews for them. As noted above, in the case of my own course structure, this piece of critical writing is produced in the preceding unit and is accordingly ready to hand. The assignment is for the student to read and assess the review generated for their text, making changes according to best practices while also providing a response to the review. This is then submitted for grading back to the instructor, who can serve as an “editor” that receives the changed “manuscript,” the review, and the student reply to the review. Taken together, this all creates a complete faux peer review process, complete with submission, revisions, and a judging editor.

The instructor is able to leverage the LLM to make all of this feasible, with the model doing most of the heavy lifting. However, it should be kept in mind that the instructor still needs to teach the unit. They need to explain the peer review process, discuss things like submission systems, and communicate the best practices for replying to reviewers, among other things, as noted above. As far as implementing the exercise is concerned, the instructor still needs to look over each review to ensure that they are meeting all of the set parameters – that is, the instructor still needs to do basic due diligence in preparing material that students are meant to thoughtfully engage with. More importantly, however, the instructor also needs to assess the student revisions, considering the changes they made, whether their response was appropriate, etc. This is also an opportunity to give individualized feedback to each student, with each assessment allowing the instructor to note things ranging from the adequacy of some change to the tone of the response.

With that in mind, we can now turn to the procedure of generating these reviews. The two tools suggested here are an e-learning platform and ChatGPT. The e-learning platform serves two purposes.² First, it aggregates all of the texts to be processed in one place with one format,

¹ For a recent systematic review of AI in higher education, see Crompton and Burke (2023). For a discussion of using LLMs to prepare reviews, including some ethical concerns, see Hosseini and Horbach (2023). For a study exploring the use of LLMs to support peer review, see Checco et al. (2021). Kousha and Thelwall (2024) present a review of recent developments relating to AI in various aspects of the publishing process. The present work does not make any statement on the use of LLMs in the peer review process outside of the classroom. It goes without saying that LLMs as such ought to be discussed in any current academic writing course, however, that goes beyond the scope of the proposed exercise.

² Examples of e-learning platforms include Moodle, Canvas LMS, and Microsoft Teams.

which makes producing a series of reviews for a few dozen essays easier. Secondly, and more importantly, it should ensure that the submitted texts are in a digital, editable format, allowing the text to be extracted by, e.g., copying it. Using an e-learning platform is not strictly necessary, but it is convenient and generally provides a good framework for the course as a whole. ChatGPT of course serves the purpose of actually generating the review. It goes without saying that one may opt for a different LLM depending on one's preferences.

The reviews themselves are generated with a prompt detailing the task and the parameters. The instructor can adjust the prompt to their preferences, with the following being the prompt used to generate reviews for my own most recent edition of this exercise:

Can you provide me with an academic review of a text? I'd like it to be around 300 words, use fairly simple language, and include two suggestions of how the text could be improved. If so, I can paste the text here.

The question structure helps the LLM understand not to immediately start inventing a review, instead having it respond that it can do that and inviting you to paste the text. The text can be pasted, with differences in formatting or spacing appearing to make no difference to the output. After it generates a review, it can be adjusted (regenerated) by asking it to be more or less critical, or sometimes including a specific issue to note in the review. Variations can also be introduced by asking it to regenerate the review but with a different introduction, etc. Once finished with a given review, it is enough to ask it whether it can produce another review following the same parameters if you paste another text (e.g., literally, "Can you produce another review using the same parameters for a different text?"). Again, this prompts the LLM to respond that it can do that with an invitation to paste the next text, which is useful as otherwise it can get confused and produce another review of the preceding text. There are other, more advanced ways of going about this, but the above is the most reliable (and the simplest) approach, in my experience.³

To this point, the system has managed to generate remarkably coherent and well-put reviews for everything it has been sent, including very odd arguments concerning things like video games or the definition of soup. What is particularly interesting here is that the LLM is sometimes able to engage with some of these topics better than I myself could. For instance, in addressing a paper about a video game, the review might include references to aspects of the game that are not described in the input text. The immediate concern is that these extensions may be hallucinations on the part of the model, but amusingly this is more of a feature than a problem from the perspective of the assignment, as addressing problematic points in reviews is part of the unit. Whereas I may not be in a position to recognize that the LLM is hallucinating

³ For those that are interested in something more advanced, ChatGPT allows for file upload, analysis, and output as of November of 2023 with ChatGPT-4. It is possible to give ChatGPT the review parameters, to upload the input essays as files, and to have ChatGPT output the reviews as files. This could save some time, particularly when processing large numbers of essays, at least in theory. In practice, having tried to do so, this has cost me far more time than serially copying and pasting texts, as suggested above. The file analysis would often fail, the reviews would repeat certain fragments in each file, etc. That being said, this may be improved in future versions of the model. Note, this file-based approach uses more resources to process and thus requires a paid account.

in the review, the student, who is obviously familiar with the material, may identify these issues and address them.⁴

A pair of examples are presented below, featuring original inputs and the LLM-generated reviews for them. The inputs are student essays that have been anonymized and recreated here with their permission. Each is followed with a brief comment. The essays are presented exactly as they were written, with infelicities and all. It should be kept in mind that these are student essays and should be read charitably. The first is a complex but fairly traditional topic; the second is more unusual and presents a nice example of how ChatGPT can handle something less expected.

Example 1

Why do we need intersectional feminism?

Intersectionality allows to think that discrimination can happen on the basis of several different factors at the same time, and we need to have a language and an ability in order to see it and address it. Intersectional feminism recognizes the overlapping and interdependent systems of discrimination based on gender, race, sexuality, class, and other identities. I shall discuss that this type of feminism is needed in our contemporary society.

The term intersectional feminism was coined in 1989 by American lawyer, civil rights advocate, and philosopher Kimberlé W. Crenshaw. It describes the overlapping of different systems of oppression. Each individual intersection depends on a multitude of different factors: gender, race, sexual orientation, religion, ethnicity, class, economic status, age, ability, (mental) health, education, etc.; and intersectional feminism acknowledges that women do not experience gender-based oppression in isolation. For example, when in the US black man got the right to vote, this did not empower black women; similarly, when women got the right to vote, black women were not included. Structural intersectionality is the collision of two or more overlapping dynamic of oppression.

Traditional feminism has historically marginalized certain groups: intersectional feminism must be distinguished from white feminism, that is led by and focused on the experiences of white, middle-class, cisgender women, overlooking the specific struggles of women from marginalized communities, such as women of color, LGBTQ+ women, disabled women, and women from lower socioeconomic backgrounds. Intersectional feminism aims to amplify the voices and address the concerns of these underrepresented groups.

By examining how different forms of oppression intersect and reinforce each other, intersectional feminism sheds light on the complex and interconnected systems of power, privilege, and marginalization that shape women's lived experiences. This understanding is crucial for developing comprehensive and inclusive solutions to address gender-based inequalities.

⁴ Furthermore, perhaps less amusingly, this is not an unrealistic issue in the publication process, as I am sure we have all gotten reviews that have left us scratching our heads. Obviously, there is a limit to the positive import of a hallucinating LLM, but to this point I have not myself seen, or received feedback to the effect that, a review was completely nonsensical.

Intersectional feminism promotes a more inclusive and effective feminist movement by recognizing the diverse needs and perspectives of women from various backgrounds. It ensures that advocacy efforts and policy solutions address the specific challenges faced by different groups of women, rather than adopting a one-size-fits-all approach.

Suppose that a woman calls herself a feminist and only fights for white, middle-class, cisgender women. Since feminism is the belief in social, economic, and political equality of every individual, despite of their sex and gender, advocating only for a restricted group of privileged people cannot be called feminism.

In summary, intersectional feminism is needed to fully understand and address the multifaceted nature of gender-based oppression, to amplify the voices and experiences of marginalized women, to challenge systemic inequalities, and to build a more inclusive and effective feminist movement that leaves no woman behind.

LLM-Generated Review:

This essay provides an insightful exploration of intersectional feminism, a concept first introduced by Kimberlé W. Crenshaw in 1989. The author argues for the necessity of intersectional feminism in addressing the multifaceted and overlapping forms of discrimination that affect individuals based on their gender, race, sexuality, class, and other identities. The primary strength of the essay is its clear and thorough explanation of how different systems of oppression interconnect and affect women's experiences differently.

The essay effectively highlights the limitations of traditional feminism, particularly its tendency to focus on the experiences of white, middle-class, cisgender women, thus excluding or marginalizing other groups. By contrasting this with the aims of intersectional feminism—to recognize and address the varied and specific challenges faced by women from diverse backgrounds—the author strengthens the argument for its contemporary relevance. The discussion on structural intersectionality and examples of historical exclusions, such as the voting rights movement in the U.S., provide powerful illustrations of the essay's main points.

However, there are opportunities to enhance the essay's impact. First, while the essay discusses the importance of intersectional feminism, it could further benefit from more concrete examples of how intersectional approaches have been or can be practically applied in feminist advocacy or policy-making today. This would provide a stronger connection between theory and practice, demonstrating the real-world implications of adopting an intersectional perspective.

Second, the essay could improve its persuasiveness by addressing potential criticisms or limitations of intersectional feminism more directly. For instance, discussing common misconceptions or challenges in implementing intersectional approaches within feminist movements could provide a more balanced view and demonstrate a deeper understanding of the subject matter.

Overall, the essay provides a compelling case for the necessity of intersectional feminism in contemporary society. With a few enhancements, it could offer a more robust discussion that

not only advocates for intersectional feminism but also shows how it can be effectively implemented to address systemic inequalities. (OpenAI, 2024a)

The review was generated following the procedure outlined above. The paper concerns a complex topic, but one that explores a fairly familiar area. The LLM hits the relevant points one would expect from a review – that is, a brief reconstruction demonstrating how the reviewer understood the text along with a few suggestions. The suggestions are reasonable and potentially actionable, with the student now having the opportunity to either implement them or decline, explaining themselves one way or another in their response to the review.

This is a good example of a standard “artificial review” for these essays. In this case, it does not challenge the student too much, while still providing them with a wide range of possible responses and modifications, inviting some reflection and potentially a bit of additional research.

Example 2

Handshakes Are an Unreasonable Social Norm

The most commonly accepted form of greeting for centuries is a handshake. Historians claim that the first notions of this form of acknowledgment date as far back as the 5th century BC. The creation of this practice might have originated as an act of showing that a person who is reaching their hand is not holding a weapon. Whereas that could have been useful in the distant past, now the custom of a handshake seems unreasonable upon closer examination. My aim in this essay is to show how handshakes are nothing but a social norm that lacks a logical component of its superior usage to other forms of greeting.

In most cultures, refusing a handshake is perceived as a form of rudeness towards the person offering it. But there are many situations where it is reasonable not to shake somebody's hand that are not often considered by many while judging others' social behavior. One of the frequently forgotten points is mysophobia - also known as germophobia. It can be defined as a pathological fear of germs and contamination. A person who suffers from this disease will usually try to avoid shaking another's hand at all cost. If the germaphobe does not present a thorough explanation of their behavior, the rejected person might feel extremely disrespected and possibly create negative feelings towards the other person. This shows how the lack of consideration from the blind acceptance of a handshake as a norm can create misunderstandings and disputes between acquaintances.

Not in all social situations people are expected to greet each other with a handshake. There are many variables regarding this issue (e.g. how well you know someone, what social environment you are in etc.) and one of them is the number of people you are supposed to greet. The issue lies in the fact that there is no specific number that everyone would consider as too big for one person to welcome with a handshake. Nevertheless, we cannot say that there is none.

Let's consider this scenario: suppose that you walk into a room of 20 people who all turn around in your direction. You have the task of greeting all of them. You might either: Say “hello” and make brief eye contact with the people in the room or walk up individually to each person and shake their hand. Now let's suppose that you go into another room where there are 3 people and you are facing the same task. The same situation applies in the next rooms where

there are 7 and 15 people. The question is whether you would choose the handshake as the form of greeting in all of those situations and if your answers vary based on how many people were present in a room.

The lack of a concrete number that would be socially considered too big to greet everyone with a handshake may lead to confusion and lack of certainty on behalf of the person greeting others as they might not be sure what kind of approach they should take.

Some might argue that handshakes are not a social norm, but a genuinely preferred form of greeting and that is why it is so popular in many cultures worldwide. It has been observed that many people welcome each other with a handshake rather than a nod or a verbal acknowledgment. What distinguishes a handshake from the other popular forms of greeting is the physical aspect of grabbing someone else's hand. So let's suppose that people generally prefer handshakes over non-physical forms of greeting (e.g.: Nodding, bowing, verbal welcome). That would mean that people prefer forms of greeting that have a physical aspect - such as hugging, over the forms that are not physical. But this argument can hardly be found in real-life examples; Usually, people do not hug others as a form of greeting (in their place of work, at a conference, when meeting someone for the first time, etc.). Therefore, people generally do not prefer handshakes over non-physical forms of greeting. This concludes that the only reasonable explanation for this phenomenon is that handshakes are a social norm and not a genuinely preferred form of greeting.

The custom of a handshake does not possess any characteristics that would rationally necessitate its popularity. The lack of clear rules regarding this specific form of greeting as well as the pressure that the general public applies on an individual may lead to many embarrassing and unpleasant encounters. Despite the existence of many reasons that give footing to an argument that handshakes are not the best form of welcoming a person, the custom of it still persists because of its social construct.

LLM-Generated Review:

This essay presents a critical examination of the handshake as a widely accepted social norm and argues against its logical necessity in modern society. The author delves into the historical origins of handshakes, suggesting that they originated as a means to demonstrate that a person was unarmed. The essay then critiques the continued use of this greeting form, highlighting its potential to cause misunderstandings and social discomfort, particularly for individuals with mysophobia.

One of the essay's strengths lies in its exploration of how handshakes can lead to social friction. The discussion of mysophobia is particularly insightful, as it highlights an often-overlooked reason why someone might refuse a handshake, thus emphasizing the need for greater empathy and understanding in social interactions. Additionally, the essay effectively questions the practicality of handshakes in various social scenarios, such as large gatherings, where other forms of greeting might be more appropriate and less cumbersome.

However, there are areas where the essay could be improved. First, the argument could benefit from a more structured and coherent flow. The points regarding the historical context, the psychological implications of refusing a handshake, and the practicality in different social settings are all valid but could be better organized to create a more compelling narrative. A clearer thesis statement at the beginning and topic sentences at the start of each paragraph would enhance the readability and impact of the argument.

Secondly, while the essay convincingly argues that handshakes are a social norm rather than a preferred form of greeting, it could be strengthened by providing alternative forms of greeting that might be more logical or suitable in today's context. For instance, discussing the growing acceptance of fist bumps, nods, or even verbal greetings could offer a constructive outlook on how social norms around greetings might evolve.

Overall, the essay provides a thoughtful critique of handshakes as a social norm, but it would benefit from a more organized presentation and the inclusion of alternative greeting methods to support its argument. (OpenAI, 2024b)

This review was also generated following the procedure outlined in this section. The subject of the paper is unusual but a good example of the kinds of papers received in my classroom, as I encourage odd topics. The LLM has no trouble, again generating what one would expect from a simple review. The suggestions are more interesting here. The first suggestion in particular can be singled out. It seems to be off, in that the paper appears to be fairly well structured, and the thesis does seem to be quite clear in the introductory paragraph.

The student needs to make a decision about what to do in this case – whether that means reworking the structure of the paper, making small adjustments, or outright declining to make any changes. The instructor then has the opportunity to address what the student ultimately did when grading the response, whether that means telling them that major revisions were not really needed, applauding them for making some small superficial changes, or whatever the situation calls for.

3. Student reception in numbers

In closing, and with a clearer picture of what this exercise consists in, we can turn to a brief and more empirically informed discussion of how it has been received by students. To that end, we can leverage two sets of data. One set is from a questionnaire that was distributed in the 2023/2024 academic year as part of a longer-term, ongoing study exploring various aspects of student engagement with AI in the classroom. The other source is annual course evaluations, which provide a few general points that may be indicative of the contribution of this assignment, though with a number of significant limitations.

The first data set is drawn from an anonymous questionnaire that students were invited to fill out after the end of the course. The questionnaire was made available to both cognitive science students and philosophy students, with the data being combined below. The questionnaire concerns student perception of the LLM-generated review and includes a few questions that have been singled out here as relevant to assessing the exercise. Three of these have been selected and presented below. There is of course a limitation here in the sample size ($n=30$), with the questionnaire being voluntary and online. However, for the purposes of illustrating the reception of the artificial reviews and the assignment, this should be sufficient.

The first two questions concern the impression the students had of the “reviews.” The first asks “Was the review easy to read and understand?” (Figure 1), and the second asks “How intuitive was the structure of the review?” (Figure 2). Each was answered on a scale from 1 to 5, with the framing language of the scales provided in the description of the figures below. As can be seen in the figures, the students did not appear to have any trouble engaging with the material the LLM produced, which is not a given in a classroom context.

[Figures removed]

[Figures removed]

Figure 1 – “Was the review easy to read and understand?” (1 –very difficult, 5 –very easy)

Figure 2 - “How intuitive was the structure of the review?” (1 - very unintuitive, 5 - very intuitive)

The third question is more direct. It reads as follows: “The review was personalized for each student - so each student received unique feedback. Did you find that interesting or more engaging?” The answer was binary, with the options reading as follows: “Yes, it was more interesting/engaging for me; Not really, it didn’t make much difference for me.” The result was 29 responses for “Yes, ...” and 1 response for “Not really, ...”. Though this is a binary choice and a fairly small sample size, the one-sided result largely speaks for itself. The anonymity of the questionnaire also gives this more credence than any feedback that might be obtained in conversation with students. Taken together, the above is indicative of a positive student reception of the exercise.

The second data set is annual course evaluations, which are also anonymous. There are a number of limitations here, however. First, course evaluations are voluntary and found online,

and so the number of respondents is relatively low (below 30% of course participants). Second, perhaps more significantly, the course evaluations are fairly general, and so it is difficult to extrapolate anything about any specific part of the course. The tables below present a selection of questions that are plausibly affected by the course design and particular exercises.⁵

Scale from 1 to 6, No framing language	Number of resp.	Course average	Faculty Average
The manner in which the classes were conducted was conducive to participation and engagement	15	5.93/6.00	5.42/6.00
Classes in the assessed subject have improved my knowledge and skills	15	6.00/6.00	5.34/6.00
I am happy I participated in this class	15	6.00/6.00	5.33/6.00

Table 1 - Cognitive science students, 3rd year, 2023 summer semester course evaluations

Scale from -2 to 2, Relativized framing language	Number of resp.	Course average	Faculty Average
The manner in which the classes were conducted was - uninteresting/interesting	14	2.00/2.00	1.54/2.00
The teaching method was not conducive to learning/was conducive to learning.	16	1.82/2.00	1.35/2.00
Please indicate to what extent you agree with the following statement: The class improved my competence in the area covered by the course	16	2.00/2.00	1.54/2.00

Table 2 - Philosophy students, 1st year, combined 2022 winter semester and 2023 winter semester course evaluations

Though only indirect evidence for the import of the assignment, the positive evaluations for a technical skills course like Academic Writing is indicative of a successful course design, which includes the exercise at issue. Whereas my own impression of the student reception of the whole faux peer review exercise has been positive, the above empirical data provides a more reliable perspective – and it is likewise positive.

3. Conclusion

The simple procedure outlined here allows for reviews like those presented above to be provided to each and every student of the course, generating compelling material for the students to address. The contribution of this work is precisely to demonstrate not just the potential of artificial intelligence for the classroom or to present the results of some specific implementation, but to present a specific application that teachers can easily implement in their own courses today. There are practically no barriers to entry in adopting it, with no imposition

⁵ The most recent 2024 semester course evaluations are not yet available to faculty and so could not be included here.

relating to institutional access to licenses or tools and nothing outside of the most basic competences required of the teacher.

Reference List:

- Bailey, S. (2021). *Academic writing for university students*. Routledge.
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, 8(1), 1–11.
<https://doi.org/10.1057/s41599-020-00703-8>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Dale, R., & Viethen, J. (2021). The automated writing assistance landscape in 2021. *Natural Language Engineering*, 27(4), 511–518.
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24.
- Hosseini, M., & Horbach, S. P. J. M. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*, 8(1), 4.
<https://doi.org/10.1186/s41073-023-00133-5>
- Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of Generative Chatbots in Higher Education. *Information*, 14(9), Article 9. <https://doi.org/10.3390/info14090492>

Kousha, K., & Thelwall, M. (2024). Artificial intelligence to support publishing and peer review: A summary and review. *Learned Publishing*, 37(1), 4–12.

<https://doi.org/10.1002/leap.1570>

Martinich, A. (2016). *Philosophical writing: An introduction*. Wiley-Blackwell.

Matzakos, N., Doukakis, S., & Moundridou, M. (2023). Learning Mathematics with Large Language Models: A Comparative Study with Computer Algebra Systems and Other Tools. *International Journal of Emerging Technologies in Learning (iJET)*, 18(20), 51–71. <https://doi.org/10.3991/ijet.v18i20.42979>

Mouser, R. (2024). Writing with ChatGPT. *Teaching Philosophy*.

<https://doi.org/10.5840/teachphil2024429197>

OpenAI. (2024a). *ChatGPT 4*. <https://chat.openai.com/>.

OpenAI. (2024b). *ChatGPT 4o*. <https://chat.openai.com/>.

Seufert, S., Burkhard, M., & Handschuh, S. (2021). Fostering Students' Academic Writing Skills: Feedback Model for an AI-enabled Support Environment. In T. Bastiaens (Ed.), *Proceedings of Innovate Learning Summit 2021* (pp. 49–58). Association for the Advancement of Computing in Education (AACE).

Woodworth, J., & Barkaoui, K. (2020). Perspectives on Using Automated Writing Evaluation Systems to Provide Written Corrective Feedback in the ESL Classroom. *TESL Canada Journal Revue*, 37(2), 234–247.