



“We’re in this together”: responding to student concerns about large language models in higher education.

Raquel Coelho, Johan Alme, Helene Asbjørnsen, Anne E. Bjune, Sehoya Cotner, Helge Dahle, Ståle Ellingsen, David Grellscheid, Antonella Zanna Munthe-Kaas, Magnus Svendsen Nerheim, Belinda Munoz Solheim, Willem Godert Maria van der Bilt

This report was generated during a workshop sponsored by the bioCEED Center for Excellence in Biology Education and SLATE, both at the University of Bergen, Norway.

SLATE Research Report 2024-1

“We’re in this together”: responding to student concerns about large language models in higher education.

© Raquel Coelho^a, Johan Alme^b, Helene Asbjørnsen, Anne E. Bjune^d, Sehoya Cotner^d, Helge Dahle^e, Ståle Ellingsen^d, David Grellscheid^f, Antonella Zanna Munthe-Kaas^e, Magnus Svendsen Nerheim^g, Belinda Munoz Solheim^h, Willem Godert Maria van der Biltⁱ

^aCentre for the Science of Learning & Technology (SLATE), University of Bergen

^bDepartment of Physics and Technology, University of Bergen

^cGeophysical Institute, University of Bergen

^dDepartment of Biological Sciences, University of Bergen

^eDepartment of Mathematics, University of Bergen

^fDepartment of Informatics, University of Bergen

^gUiB Learning lab, IT-department & Portfolio-board, University of Bergen

^hDepartment of Education, University of Bergen

ⁱDepartment of Earth Science, University of Bergen

Contact Information:

Sehoya Cotner (Sehoya.Cotner@uib.no)

bioCEED Centre for Excellence in Biology Education

Department of Biological Sciences

University of Bergen

Postbox 7807

5020 Bergen

ISBN: 978-82-93789-10-9

To cite this report:

Alme, J., Asbjørnsen, H., Bjune, A. E., Coelho, R., Cotner, S., Dahle, H., Ellingsen, S., Grellscheid, D., Munthe-Kaas, A. Z., Nerheim, M. S., Solheim, B. M., van der Bilt, W. G. M. (2024). “We’re in this together”:

responding to student concerns about large language models in higher education. SLATE Research Report 2024-1, Bergen, Norway: Centre for the Science of Learning & Technology (SLATE). ISBN: 978-82-93789-10-9

ChatGPT Plus with DALL-E 3 was used to generate the cover image.

Table of Contents

Table of Contents	3
Abstract/Executive summary	4
About us	4
Student surveys	5
What do the students say?	5
Large Language Models as educational tools.....	5
When are LLMs useful for the learning process?	5
What should they not be used for?	5
Can I trust LLMs?	6
Acceptable use, fairness, and ethics.....	6
General concerns about fairness.....	6
Concerns about ethics and whether the use of LLMs is unethical	6
Concerns about lack of transparency from the instructor on how/if these tools should be used.....	7
What are the boundaries for acceptable use vs cheating?.....	7
What about assessment? In what context is it okay to use LLMs in assessment situations?7	
Do other students have an advantage if I don't use it? Will lack of competence with LLMs put me at a disadvantage?	7
How can we address student concerns?	8
Communicating with our students	10
Conclusion/Next steps	12

Abstract/Executive summary

We, the self-proclaimed “LLMs¹ in Education” group, comprising university educators (and one partner researcher) from the University of Bergen, Norway, conducted two workshops in August and December 2023 to explore the use of tools powered by large language models (LLMs) in education. We analyzed student responses from two surveys, one national (n=660) and one local (n=26), to understand their perceptions and concerns related to tools such as ChatGPT. Specifically, we focused on responses to the question “How do you think these tools should be used in education, in a way that is fair and that supports your learning?”

We categorized the concerns that emerged in our reading of this student feedback, and we were able to group these categories into two broad themes: questions and concerns about the tools themselves, and questions and concerns related to acceptable use, fairness, and ethics. We provide our collaborative responses to the identified questions and concerns in both categories.

In sum, we advocate for honesty, transparency, and working *with* students as we navigate these new tools in teaching and learning. Students are definitely looking to their instructors to identify boundaries and best practices for LLMs in our courses, but as with many new technological advances, we may assume that our students are far beyond us in terms of savvy use of these tools. This is where a continued dialogue will be helpful in the months ahead. We will continue to ask our students about their use and perceptions of LLMs, etc., and to be fair, we should be honest with them about our own practices, confusion, and concerns.

About us

The report presents the results of a discussion on 5 December 2023, specifically involving university educators from several mathematics and natural sciences disciplines at the University of Bergen, Norway. Our group has been focused on the use of tools powered by large language models (LLMs), like ChatGPT, in education, and is therefore called the “LLMs in Education” group. Our activities have been anchored on two workshops. In our initial meeting, in August 2023, we convened to discuss what we knew about these tools, identify challenges and opportunities, and create implementation plans for using—and evaluating the use of—LLMs in our courses. Our second meeting, in December 2023, was intended to share experiences from our recent use of LLMs in our courses, and to discuss next steps based on these experiences. During the meeting, we also considered data from some recent surveys, in which students voiced concerns about the use of LLMs in their education. Ultimately, we decided that it was our duty to give thoughtful responses from our perspectives as course leaders to these student concerns.

¹ Acknowledging that large language models (LLMs) and the tools they power are not the same and recognizing that generative AI tools include more than just text generators, we use their names interchangeably in this report.

Student surveys

We report here on student responses from two different surveys: 1) a national survey (with 660 responses) of biology-related study programs across nine Norwegian universities (several items mirror those asked of students at the University of Bergen in February 2023 (reported in Møgelvang et al., 2023); and 2) a small survey (with 26 responses) of students currently enrolled in courses in which LLMs have been used (specifically, courses by one of the authors of this report). While students were asked several questions about their use and perceptions of LLMs, in this report we focus on student responses to one item: **“How do you think these tools should be used in education, in a way that is fair and that supports your learning?”**

What do the students say?

Student responses to the question “How should these tools (LLMs) be used in education, in a way that is fair and that supports your learning?” mostly reflect student confusion and concern. In fact, few responses specifically mentioned ways in which the respondent would like LLMs to be used in their education. Several categories emerged in our reading of this student feedback, and we were able to group these into two broad themes: 1) questions about the tools themselves, and 2) questions related to acceptable use, fairness, and ethics. These themes and categories, along with sample student responses, are presented in Table 1.

Table 1. Themes, categories, and examples reflecting student responses to “How should these tools (LLMs) be used in education, in a way that is fair and supports your learning?”

Large Language Models as educational tools	
When are LLMs useful for the learning process?	<p>“I think they can be useful in short term ways, like finding a good definition or providing an alternative explanation. But I think they should not be used on a grander scale, as it is important for learning to write/ think about the topics yourself.”</p> <p>“These tools should be able to be used to check whether what I have come up with is correct or not. I should be able to use LLMs to find out if my results, in e.g. a math problem, are correct, without using it to calculate the whole problem for me.”</p> <p>“LLMs can be used to find good sources for assignments if they only refer to research articles or the like. They can perhaps also be used for summaries of articles, or as a kind of assistant teacher by answering simpler questions.”</p>
What should they not be used for?	<p>“People should learn about them and learn how they work. they should be shown that the LLMs are not always correct so that people know not to trust blindly in them. it might be good to also be sure to teach students how to find much of the same type of information by use of other search engines etc.”</p>

	<p>"Now many LLMs are used as a helper so that you don't have to think for yourself. One does not have to think through so much anymore and this gives the impression of having the potential to reduce the depth of understanding among students. By actively using it in teaching, one could perhaps show how LLMs can be used for learning in this subject and what is not a good idea to do."</p> <p>"I think it does not support learning and is fair when students, for example, use chatGTP to write text for themselves/do the work for themselves. Then I would like to think that there will be less reflection, and you don't seem to think or write yourself."</p> <p>"I think it makes me lazier, I'd rather ask chatGPT for an answer than find out where I can find the answer on my own."</p>
<p>Can I trust LLMs?</p>	<p>"I think LLMs should be able to be used as a tool to find ways to solve problems in simple mathematics and statistics subjects, but under no circumstances to write texts. In my experience, you cannot trust the sources the programs use. In addition, they can get information wrong. When it comes to longer texts you do yourself a disservice by using the programs, as you don't get to practice writing, being critical of and interpreting information."</p> <p>"I don't think they should be used at all except to show the dangers of AI generated sources for example."</p>
<p>Acceptable use, fairness, and ethics</p>	
<p>General concerns about fairness</p>	<p>"I think it does not support learning and is not fair when students, for example, use chatGTP to write text for them or do the work for them. Then I think that there will be less reflection, and you neither think nor write for yourself."</p> <p>"It'll give an unfair advantage to those students choosing to use it by themselves over students who wrote it all themselves."</p>
<p>Concerns about ethics and whether the use of LLMs is unethical</p>	<p>"AI can only end poorly for us all. It will just blur the lines of who did what, encourage people to lose artistic and creative skills, and make people even lazier than they are now. Ban it now before it's too late."</p> <p>"The educational institutions should background check the ethics of LLMs, advocate for ethical use/creation. make sure the AI used/suggested isn't made/filtered by people sustaining modern slavery etc. and advocate for ethical laws being enforced around the use and creation of AI."</p> <p>"I think it should be used ... primarily to assess and improve texts I have produced myself. I think there is an ethical limit when such methods are used to produce a primary text."</p> <p>"I do not think this has any ethical implication."</p>

<p>Concerns about the lack of transparency from the instructor on how/if these tools should be used</p>	<p>"I would really appreciate getting a little course on how to use these tools efficiently and that we get to know how it's important to be critical towards the information we get from these tools."</p> <p>"There should be training in the use of and encouragement to learn how to use these tools. As well as clear guidelines for when and how it should be used."</p>
<p>What are the boundaries for acceptable use vs cheating?</p>	<p>"It would be great if teachers would use LLMs and give us some confidence that what we are using is okay."</p> <p>"Would have liked some training in how to use these tools without falling into the trap of cheating or incomplete work."</p> <p>"As long as LLMs can be checked for plagiarism, it can be fine, but I think it's an easy way to "cheat"."</p> <p>"Since they are still relatively new, correct use of these tools should be discussed and encouraged in class."</p>
<p>What about assessment? In what context is it okay to use LLMs in assessment situations?</p>	<p>"Perhaps by emphasizing and recognizing that there are good aids for the students. ChatGPT is e.g. very good at getting a concrete experimental setup to answer a research question. By the lecturer recognizing this, that the students have the opportunity to find out with the help of a simple search, can ensure that such questions are asked more often in formative assessments and create perspective. Whereas in cases of exams, one can search for understanding and creative problem solving instead. I think this is only positive and provides greater learning than a specific solution being devised for a specific research question."</p> <p>"I think there should be term papers (with LLMs) + exams (without LLMs) in all subjects where this is possible."</p>
<p>Do other students have an advantage if I don't use it? Will lack of competence with LLMs put me at a disadvantage?</p>	<p>"I never blindly trust that what the bot says is 100% right. It is, however, a good public tool that is used by many, and one can feel that one is falling short if one does not use the chat as a tool (compared to others who use it), since in some cases one takes longer to arrive at similar results on their own."</p> <p>"I feel like if it's not mentioned or given guidance on it, it'll give an advantage to those students choosing to use it by themselves in a way giving them an unfair advantage over students who wrote it all themselves"</p>

How can we address student concerns?

In considering these general concerns, we discussed honest and constructive feedback. The responses to many of the topics raised by the students are context-dependent, and likely to vary based on discipline, course level, specific learning objectives, and instructor preferences. Thus, in many cases we offer a range of possible responses.

When are LLMs useful for the learning process?

- As with many of these concerns, the answer will depend on the course learning objectives. If generating a testable hypothesis is a learning goal, then the students should not start by asking ChatGPT to generate a testable hypothesis. However, an LLM can give feedback after the fact on something a student has generated or present them with a range of other options to help with self-evaluation.
- They are useful as a study pal, i.e., for making summaries, quizzes and explaining complex concepts, but this is also dependent on the student's knowledge and understanding of the subject, as these are crucial for critically engaging with the generated texts.
- Interacting with an LLM when writing code can be useful to get 'unstuck.'
- They are good for improving language, doing translations, etc.

What should they **not** be used for?

- LLM-powered tools should not be used as a search engine. We see many of our students using them as search engines. Most of these tools are not designed for search purposes, as they generate text based on statistical patterns of word usage, which can lead to inaccurate or false information. When conducting a search, it is crucial to access links to reputable sources to ensure the reliability of the information.
- Using them for fact checking is right now risky, as they are not reliable (and might never be).
- They are not creative, and they cannot reflect. Keep in mind that the output they generate should not be interpreted as creative or reflective. By definition they are producing text based on likelihoods of association with prior text—they cannot, and will not, generate a novel idea.

Can I trust LLMs?

- You can trust LLMs to generate coherent text. But looks can be misleading! Just because a text reads well does not mean that it is not full of misinformation, irrelevant information, or outright fabrication.
- Be sure to check all citations generated by LLMs. These are usually made up, or irrelevant to the content.
- Note that the information generated by an LLM will reflect the inherent biases and misconceptions present in the source data. ChatGPT, for example, is no less sexist or racist than the massive sources that have been used to train the model.

General concerns about fairness

- Each instructor will have different rules and expectations. If these aren't clear, ask! Each instructor is responsible for communicating the same set of expectations to the entire class.
- Several of your instructors use rubrics or may use this opportunity to develop rubrics for your assignments. One of the reasons we like rubrics is to avoid bias in grading, based on, for example, an especially well written text. A careful rubric can force us to look

beyond good writing to evaluate other aspects of the assignment, thus avoiding the “halo effect” in grading (“if the writing is this good, everything is probably good”).

Concerns about lack of transparency from the instructor on how/if these tools should be used

- Instructors are also in a “learning phase”, i.e. many instructors have yet not decided how or if to use these tools in their teaching. Ask for an open discussion about their use.

Concerns about ethics and whether the use of LLMs is unethical

- We have discussed this quite a bit, and it is reasonable to express ethical concerns. For example, there are large environmental concerns: one interaction with an LLM is orders of magnitude more resource-demanding than a conventional search engine interaction, regardless of the upfront cost of training the model. There are also concerns about LLM results that, due to biases in training data, simply mirror the built-in social biases of the last decades: sexism, racism, etc. There are also copyright and academic integrity issues, as these tools are constructed from unacknowledged source material. Further, there are ethical concerns about the use of extremely low wage workers to train the models.
- We also realize that problems of representation in outcomes cannot be ignored by saying it works for the majority of users. The developers of LLMs (those who have more of the resources) place the hard work of verification and bias identification on the end users. We cannot resolve this problem, but we can be aware of it, and use a critical lens to question the existence of bias in LLM output.
- One challenge is that some models are better than others, and perhaps the best ones are the most expensive ones: the more you can afford, the better the tool you can access. Thus, LLMs could exacerbate existing inequalities.
- However, LLM use could help “level the playing field” for some students, e.g., those working in a second language, or those with communication barriers such as dyslexia.

What are the boundaries for acceptable use vs cheating?

- We think it is helpful to reframe this in terms of *responsible versus irresponsible use*. And, again, instructors will vary in what they consider acceptable use. What is considered acceptable to achieve some learning objectives may be unacceptable to achieve others.
- Using it to improve text is OK (if used critically) - using it to generate text, and to then represent that text as your own, is never ok.
- Basically, it is ok if it is being used to improve your own understanding (e.g., as a Socratic tutor); however, you are ultimately responsible for curating what the LLM generates and ensuring it represents your own understanding.
- The take-home message is that you should ask your instructors for clarification on what they consider acceptable use of LLMs.
- There are also national and academic discussions on these topics. Examples include resources from Common Sense Education (<https://www.commonsense.org/education/articles/chatgpt-and-beyond-how-to-handle-ai-in-schools>), the University of California Berkeley (<https://ethics.berkeley.edu/privacy/appropriate-use-chatgpt-and-similar-ai-tools>), and UNESCO (https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf).

What about assessment? In what context is it okay to use LLMs in assessment situations?

- The use of an LLM during exams or other assessment situations should be clarified with the instructor so that the same guidelines are communicated to the entire class (from a fairness perspective).
- Assessment based on quality of language in MSc theses, for example, should probably be deemphasized.
- Many of us have been discussing assessment practices during this age of LLMs. This is a great opportunity to reconsider our assessment practices to ensure that they are student-centered, evidence-based, and equitable. If an assessment is easy to complete with an LLM, maybe it is not the best assessment.
- In sum, instructors should give clear instructions on use of LLMs during assessment activities.

Do other students have an advantage if I don't use it? Alternatively, will lack of competence with LLMs put me at a disadvantage?

- In the short-term students that use LLMs might have an advantage, but in the longer term it is more beneficial to learn and understand the course material yourself without over-reliance on an LLM-powered tool.
- However, it is wise to familiarize yourself with new tools that might improve your efficiency and understanding, as you will also encounter them in your working life.
- Resources exist to help with familiarization, and may include local offerings (e.g., the DIGI courses at the University of Bergen).

Communicating with our students

Many of these student concerns can be mitigated by careful communication with their instructors. One of us (SE) has led an in-class exercise in which students discuss, in small groups, what rules the class should adopt with respect to the use of LLMs in their course assignments. This led to clear communication, discussion to clarify concerns, and a list of useful guidelines that could be shared with the class.

Several other sources (see below) have suggested ways to communicate policies on responsible use of LLM-powered technologies to students. We offer the following specific language, specifically in response to the concerns voiced by our own students.

Example 1. Tools build on Large Language Models (specifically for me, ChatGPT) have become a useful part of my own work and I understand that you may want to use them in your course assignments. I encourage you, however, to learn about how these models are created, and why what they generate is typically well written, but bland and not creative and potentially wrong. You are much more interesting than ChatGPT, and I want to read your thoughts! Further, the information generated will reflect societal biases including sexism and racism, so approach any generated text with a critical lens.

Example 2. If you generate text from ChatGPT, and then use that text verbatim as your own work, I consider that unacceptable use. Please don't do it. For some class activities, generating text to prepare (e.g., for a debate) will be fine, and I will let you know ahead of time. I may even suggest you use an LLM-powered tool.

Example 3. I imagine many will use Generative AI (GAI) tools such as ChatGPT, DALL-E, etc. to support your work in this course. Please keep in mind some ethical and quality-control considerations.

Ethical issues: Tools like ChatGPT can be useful in idea-generation (i.e., brainstorming), onboarding (i.e., summarizing concepts), and formative assessment (i.e., testing your knowledge of a subject) processes. You may also find it helps to improve the quality of your own written text. However, while I doubt this will ever be necessary, if you do decide to use language copied from a text generator, that language should be represented in quotation marks and the tool should be cited. I much prefer you do not use text copied directly from a text generator. Also, do not upload personal information about your peers, instructors, or others to an LLM. Doing so is not only a serious violation of privacy but also illegal under GDPR regulations.

Quality issues: Be aware that there are legitimate criticisms of LLM output. These tools can fabricate content and make it look legitimate. They also generate irrelevant or false citations. Finally, they are trained on content that is biased toward English language speakers, North American perspectives, and pervasive societal biases. Approach any output with a critical lens, check citations, and ensure that any material you submit reflects your own understanding and perspectives. If you use output directly from large language model tools, provide a clear indication using appropriate referencing styles.

For additional examples on how to communicate with students, please refer to the following resources:

- [Policies available by other educators](#), Eaton, L. (n.d.). *Classroom Policies for AI Generative Tools—Google Docs*.
- [Template Syllabus Statements for Generative AI Use](#), Teaching and Learning Hub, Stanford Graduate School of Business.
- [Policies for the use of AI in courses](#), Office of Undergraduate Education, Harvard University
- [APA - How to cite ChatGPT](#), American Psychological Association Blog
- [STEM LibGuide Resources: Citing ChatGPT](#), Library, University of California, Santa Barbara
- [ChatGPT and other generative AI tools](#), Library, University of Queensland
- [ChatGPT and Generative Artificial Intelligence \(AI\): AI-generated content and citation](#), Library, University of Waterloo

Conclusion/Next Steps

We recognize that these insights are speculative and may evolve as more research data and user feedback becomes available. Yet in sum, we advocate for honesty, transparency, and working *with* students as we navigate these new tools in teaching and learning. Students are looking to their instructors to identify boundaries and best practices for LLMs in our courses, but as with many new technological advances, we may assume that our students are far beyond us in terms of savvy use of these tools. This is where a continued dialogue will be helpful in the months ahead. We should have an ongoing dialogue with our students about their use and perceptions of LLMs, etc., and to be fair, we should be honest with them about our own practices, confusion, and concerns.

We referenced the following report:

Møgelvang, A., Bjune, A.E., Coelho, R., Cotner, S., Grellscheid, D. and Ellinsen, S. (2023). Initial (March 2023) uses and perceptions of chatGPT in a sample of students and instructors at the University of Bergen (UiB), Norway. SLATE Research Report 2023-1, Bergen, Norway: Centre for the Science of Learning & Technology (SLATE). ISBN: 978-82-93789-09-3