Academic teaching and AI: The central role of policies

Holm, Jacob Rubæk; Henning, Martin; Peltonen, Jaakko; Timmermans, Bram

Publication date:
2024

Document Version
Other version

Link to publication from Aalborg University

Citation for published version (APA):
Academic teaching and AI: The central role of policies

_Nordic AI-BEST Final workshop; May 2024, Bergen. Summary report and recommendations_

Jacob Rubæk Holm, Aalborg University Business School, Aalborg University. 
jrh@business.aau.dk

Martin Henning, School of Business, Economics and Law, University of Gothenburg

Jaakko Peltonen, Tampere University

Bram Timmermans, Norwegian School of Economics

Introduction

This report is the 3\textsuperscript{rd} and final report of the Nordic AI-BEST project (NPHE-2023/10475). The recommendations herein, however, cannot be considered ‘final’ as the project continues as NPHE-2024/10166, and the recommendations will continue to evolve. It would maybe be misleading to expect recommendations that are in any sense ‘final’ at all, as recommendations will continue to evolve as long as AI technology evolves. Instead, it will always be fruitful to focus on the ‘most recent’ recommendations, while being aware that recommendations will eventually become obsolete with technological change.

The Nordic AI-BEST project has also had other aims than formulating guidelines. In line with the initial stated aims, the project has encouraged the participating teachers to experiment and reflect critically. In addition, the project has, to varying degrees, inspired the development of guidelines for AI use across the participating institutions, which was another aim of the project. The inspiration has come from the participants but there has also been cross-institutional inspiration. This report makes the most recent developments available to teachers and institutions outside of the project.

Throughout this project ‘AI’ refers to generative AI. The main focus is large language models (LLMs), but the project is not restricted to LLMs. It also includes for example the use of generative AI to assist in analysis and to create illustrations for presentations, reports, etc.

Policies

The most recent recommendations for policies prior to the Final workshop in Bergen were the following:
1. Have a clear institutional level policy on what is the right and what is the wrong way to use AI.
2. Let the policy contain limitations but also suggestions and encouragement for use. User examples are helpful for students and educators.
3. The policy should contain guidelines on which generative AI to use and which to avoid.
4. Institutions should devote additional resources towards teaching development to facilitate that educators can experiment with AI in teaching.
5. The policy should contain guidelines on what educators can demand from students and what cannot be demanded.

Policies across the participating institutions have evolved significantly over the past semester, and some convergence has been observed, although the correspondence between policy and practice varies. This is likely a result of variation in the awareness of policies among students and staff across the participating institutions. We have advocated for institutional level policies (recommendation 1 above), and it has been interesting to see how policies at different levels have been rolled out across the participating institutions. It appears that awareness is hindered by a well-meaning decentralisation of more concretized and detailed policies where local units, down to individual study boards, are tasked with making own policies. This is apparently born out of an expectation that a one-size-fits-all policy will fit none. However, at some institutions, the lack of simple overall guidelines prohibits use of AI by creating insecurity regarding the restrictions for each specific setting.

Even at institutions where overall guidelines exist, a general-level permission for AI use may still leave specific details of permissible uses and best practices to be worked out in the specific settings. A solution would be for the few general bans to be formalised at the top institutional level, while subunits each formulate the positive examples of use. For example, each course description could list recommended uses of AI as a personal student tutor in the course context and refer to the overall guidelines for specific limitations.

Policy and AI literacy

There is increasing awareness among both students and teachers of the limitations of their own skills in using AI. Often, they feel that they are not able to prompt the AI in a way that the AI outputs a desirable and useful output. This may be framed as a need for better understanding of prompt engineering, but such a framing can be too superficial if the main issue is a lack of understanding of how the technology works. Some users use prompts that resemble search strings for web searches, but as put by a colleague outside the project, this is equivalent to just shouting the query in a crowd. Institutional action should expand beyond policies that sets limits for use and encourages specific uses, to also include resources that allow users to become familiar with generative AI and to build AI literacy. It is even possible that there could be some cases where the use of AI in some fashion should be mandatory, for
example as part of ensuring sufficient diligence in information finding, or as part of a suite of required manuscript quality checking steps.

**Reporting AI use as a policy**

A safeguard that has been observed at some institutions is a requirement for students to report their specific use of AI; for example, in a template with an attempted exhaustive list of AI uses. Such policies can lead to overreporting. This is so far only observed in a few cases, but it has a negative bureaucratizing effect on AI use. One problem is that it is hard to justify the need to explicitly declare AI use without requiring students to make similar, often irrelevant, declarations on the use of other tools. It is hard to see that it would make sense to declare that AI was used for specific tasks without requiring other tools and sources of support to be declared too. E.g. friends, family, and AI can all be used to generate ideas or to proofread. Inspiration for content or structure of a written product can be gained from AI or from similar outputs produced by students in previous cohorts. As this form of reporting relies on students’ self-reporting after the paper or report has been handed in, it is not clear what the practical use of the reported data may be.

There is an increasing tendency for AI to become integrated in other software, such as Copilot in MS Word. This has benefits as embedding AI in software mitigates at least two issues: i) there is no longer a need for prompting or, indeed, a deeper understanding of how the technology works. ii) use in software will typically restrict use to be within the domains of the AI so that the user has less need to understand the limitations of the AI, although embedded AI can still sometimes perform poorly within its domain hence users should not rely too heavily on it. However, embedded AI also creates a problem for policies, as it may become less and less obvious when AI is used. Use in software will be to some extent automatic and not require active agency – a direct opposite to opening a browser, logging into the preferred AI service, and prompting the AI.

An observation that underscores the importance of policy is the observation of some gender bias in the use of AI (see also the section on students’ use). It appears that policies can go a long way towards mitigating this bias, as the bias appears to be correlated with different propensities to adhere to – and rely on – explicit policies. Policies for introducing AI to students would also alleviate the problem that much introduction to AI is currently ‘guerilla’ introduction, where teachers autonomously introduce AI in their respective modules. This obviously has the risk of repeating the same introduction multiple times. There are modules where AI introduction is naturally included – e.g. thesis methodology – a coherent plan for how AI is introduced to students across semesters is nevertheless required.
Updated recommendations for policy:

1. Have a clear institutional level policy on what is the right and what is the wrong way to use AI.
2. (Changed) Policy should contain suggestions and encouragement for use. These can be formulated independently at institutional subunits.
3. (Changed) The policy should contain guidelines on which generative AI to use, which to avoid, and how to access resources for building AI literacy.
4. Institutions should devote additional resources towards teaching development to facilitate that educators can experiment with AI in teaching.
5. (Updated) The policy should contain guidelines on what educators can demand from students, what cannot be demanded, and who has the responsibility to introduce AI to students.
6. (New) Policies should be revised regularly to keep up with technological change in relation to AI, not least the integration of AI in other software.

Students’ own use as a tutor

The most recent recommendations prior to the workshop in Bergen were the following:

1. Students need to be aware that the ability to evaluate the output from an AI and decide on its practical usefulness only comes with being able to do the same task yourself. You only know if an AI-generated or AI-assisted essay on a given topic is any good if you have the skill to write an essay yourself.
2. For as long as the use of generative AI is relatively novel, students need to know both limitations and potentials of the AI. Additionally, use of AI in university-level learning contexts can have different depth of use and best practices than AI use in other contexts. The limitations and potentials can be taught through examples in class by educators, relatively structured assignments leveraging AI, and by peer learning among students.
3. Teachers should avoid training students to solve problems only with access to an AI: While AI can assist with problem-solving, it’s important to recognize that students should not rely solely on the technology to provide answers, and thus not be able to solve a problem without access to an AI.
4. Notice that competitiveness between students can lead them to be reluctant to share tips and tricks for using AI as a tutor. Incentivize sharing through e.g. peer grading and group work.
5. Explore solutions for local AI services so students do not need to create profiles with third parties, and the contents of prompts remain local.
6. Train students in ethical and responsible use of AI from the first semester. This specifically means giving them knowledge on the institutional policy on using AI and
available AI software, if any. The training could be part of an existing introductory module introducing students to other policies and facilities at the institution.

Students’ skills for using AI appear to be progressing rapidly but unevenly. They can be observed to experiment with prompting, with varying success. Those who struggle appear to have a limited understanding of how AI, and especially LLMs, work. Yet, students are also seen to leverage AI meaningfully e.g. dyslexic students using AI support for written hand-ins. In a few cases, students have also been observed to use AI in class to generate questions or comments to the teacher, which the student then presents as his/her questions or comments. The motivation appears to be a desire to appear engaged in the teaching, but it may not be the preferred form of engagement from a didactic viewpoint.

**Observed differences among students**

There is some evidence, collected outside of Nordic AI-BEST by e.g. Anja Møgelvang from the Western Norway University of Applied Sciences and, independently, by Catalina Franco from the Norwegian School of Economics, that the use of AI among students is gender biased. Put simply, students classified as female often fear the consequences of using AI and thus refrain, while students classified as male fear the consequences of not using AI and use AI even if not allowed. These are crude generalizations as the use of AI and the inclination to disregard restrictions are also found to depend on, inter alia, academic proficiency. Students that have high academic proficiency trust in their own abilities and use AI less, while students aware that their academic proficiency is less than top of the class tend to use AI more. Opting out of using AI is also observed when it is unclear to students if AI use is cheating. They then refrain from using AI or avoid making their use of AI observable to peers as well as teachers. In contrast, they have a clear and acute understanding that there is a benefit to build competencies for using AI for the post-graduation job market.

Over the spring 2024 semester it has become increasing possible to observe two different uses of AI among students: AI as an efficiency enhancing tool that can be used to cut corners and complete course work with less effort and time, and as a learning support tool that can be used to improve own learning. These two forms of use are intertwined: When a task is first encountered students can leverage AI to build the skills needed to complete the task. After these skills have been built, students can use AI as a shortcut to complete the task because the skills will allow the student to evaluate the work done by the AI. It is important to assist students in realising this distinction.

In some cases, such as written hand-ins, AI used for efficiency enhancement by students risks being classified as cheating. This is quite natural as increasing efficiency by letting AI do your assignment work in ways or extents that go beyond what is allowed by university and course policies or without sufficient reporting of its use is indeed cheating. This demonstrates
how assessment needs to be adapted to AI use by students: mastering the knowledge, skills, and competencies for a course needs to be assessed by the critical reflection that students apply when evaluating output from AI.

Unfortunately, students do not always see the line between cheating and using AI to support learning, and thus refrain from using AI in order to not risk being accused of cheating. They are equally concerned about this accusation coming from the teacher, the university, or from their peers. Cheating is, however, so far not a large problem. Or at least it is confined to a specific type of module where the incentive to cheat is high because the risk of getting caught is low. Supplementing the evaluation of written reports with an oral assessment, for example, can often reveal if the student has used AI to produce the report, but lacks the skills to critically evaluate AI output.

All in all, these, developments and observations among students do not lead to updated recommendations on students specifically. However, they point to updates on policies, as discussed earlier, and updates on assessment, as discussed below.

Teachers’ uses in class and in preparation for class

The most recent recommendations prior to the workshop in Bergen were the following:

1. Using AI to prepare teaching is a ladder of four steps. Do not get stuck at the first step: “Initial Shock”.
2. Using AI can provide a large benefit to students so ensure that it is equally accessible.
3. AI can have low legitimacy in the eyes of students and steps can be required to mend this in the context of applications where AI could be useful.
4. Moving up the four-step ladder can be furthered by a change in curriculum. What these changes should be and the degree to which they are specific to each course needs to be explored.
5. A promising venue for efficiency enhancement that can lead into new teaching solutions is to leverage the creative power of generative AI to create cases and games. These can substitute empirical cases, although the educator may then need an alternative way of demonstrating that the course material is practically relevant.
6. The role of generative AI in students’ written output, including theses, needs further experimentation and discussion.

It is being observed that the expected benefits in terms of efficiency enhancement are hard to realise. In general, using AI in class or to prepare class can be associated with self-fulfilment – i.e. be fun – but any increases in efficiency, allowing for less time spent on class preparation, or increases in student learning are hard to identify. Thus, it is not easy to recommend changes to the curriculum facilitated by AI. When implementing AI in a course it will often be
better to maintain a flexible curriculum and then let the teacher present AI use as part of the didactic contract of the course or of specific sessions. However, a side effect of teachers’ use of generative AI is that familiarity with AI makes it easier to spot AI output when handed in from students; at least in instances where the students’ use is relatively basic.

Some of the developments observed among students are also observed among teachers. AI literacy is developing rapidly but unevenly, and a significant share of colleagues could benefit from a deeper understanding of how generative AI works to further their creative use and efficient prompting. Teachers (and students) are very reflective and eager to discuss their use of AI, even when they are reluctant users.

One potential use that can take work off teachers is to construct a chatbot based on the curriculum for a course, or at least based on the teacher’s material. Students can then e.g. be recommended or even required to ask the chatbot any question regarding the module before asking the teacher. Of course, some students may have already figured this out themselves and built such chatbots. And, at least for the academic and not practical questions, it appears likely that publishers will make such chatbots available before long. Thus, it can be expected that teaching chatbot assistants will become relatively common before long, and it can potentially have consequences whether they are built by teachers, students, publishers, or even third parties.

**Why is teaching so hard to automate?**

Except for the use of AI to generate assignments, data, cases etc. for students it has proven difficult to increase efficiency for teachers. Thus, also moving towards new teaching solutions appears to be far off. AI has underscored just how personal teaching really is; even when it is old fashioned lecturing in a large auditorium. AI can condense a text into PowerPoints and narrate it faster and at least as good as a teacher, but students have always, rightfully, complained that teachers need to do more than just repeat the text that the students have already read. Rather, teachers need to contribute the reflections on theory of an experienced researcher, and perspectives on current events. Assignments, on the other hand, may well be generated by AI rather than the teacher although their quality should in many cases still be verified by the teacher. However, assignments are to some extent already provided by publishers of textbooks, and publishers are likely working on AI powered assignment banks already.

As in the section on students’ own use, these developments and observations do not lead to updated recommendations. However, they point to updates on policies, as discussed earlier, and also updates on assessment, as discussed below.
Use in assessment

The most recent recommendations for assessment prior to the workshop in Bergen were the following:

1. The most practical assessment where students cannot benefit from using AI may be the oral exam, or written exams in a controlled environment disallowing personal devices.
2. It is important to be careful if using AI to prepare an exam – including oral exams – as each student should have the same precondition for undertaking the exam. A teacher may produce different questions for different students’ oral exam – and indeed is likely to do so because students are different and respond differently – whereas the randomness introduced by letting AI prepare an exam individually for different students without effective curation by the teacher is less defensible.

So far, generative AI has had a conserving effect on exam formats where relatively new and more efficient, or at least cheaper, exam formats such as the take-home exam are being abandoned. Instead, written sit-in exams are moving back on campus. However, such exams are not impervious to cheating with generative AI as it has been observed that students will on occasion bring a large amount of AI generated output to the exam, and then copy text from the AI generated output. This can of course be mitigated with a closed book exam, but it has also been observed at oral exams where students read out loud from the AI generated notes. When assessing exam papers or other hand-ins, the style of writing has previously served partially as a signal, as high-quality writing correlated with high quality content. But this correlation does not exist anymore when AI can produce high quality writing without quality content. This has made grading more arduous as the grading teacher needs to argue carefully against very eloquent nonsense.

In cases where it was not feasible to change the exam format from take-home exams clear signs of cheating were observed. Thus, it is not surprising to observe that part of faculty across the participating institutions are still at the ‘panic’ phase of AI adoption, quickly changing all exams to sit-in exams to block AI use.

With the above limitations to assessment in mind, it is nevertheless necessary to develop exam formats that allow for the assessment of students’ knowledge, skills, and competencies when they do use AI. In most cases that are relevant in academia, AI does not make skills superfluous but creates a shortcut to imitate skills. Thus exams should evolve from assessing students’ ability to *demonstrate* skills and towards assessing students’ ability to *assess* skills themselves.
Updates recommendations:

1. The most practical assessment where students cannot benefit from using AI may be the oral exam, or written exams in a controlled environment disallowing personal devices.

2. It is important to be careful if using AI to prepare an exam – including oral exams – as each student should have the same precondition for undertaking the exam. A teacher may produce different questions for different students’ oral exam – and indeed is likely to do so because students are different and respond differently – whereas the randomness introduced by letting AI prepare an exam individually for different students without effective curation by the teacher is less defensible.

3. (New) To evaluate students’ knowledge, skills, and competencies it is no longer sufficient for students to demonstrate knowledge, skills, and competencies, they must also demonstrate the higher order competencies of evaluating knowledge, skills, and competencies.

The road ahead

The report marks the end of the first part of the Nordic AI-BEST project but the project now continues in a second phase where recommendations will continue to develop, but where project participants will also research students’ use of AI more formally.
Workshop participants

University of Tampere, Finland:
- Jaakko Peltonen
- Suvi-Päivikki Ikonen

Norwegian School of Economics, Norway:
- Bram Timmermans
- Vidya Oruganti
- Björn Schmeisser
- Catalina Franco
- Margrete Dyvik Cardona

University of Gothenburg, Sweden:
- Martin Henning
- Roman Martin
- Ramsin Yakob

Aalborg University, Denmark:
- Jacob Rubæk Holm
- Kristian Nielsen
- Pernille Gjerløv-Juel

Western Norway University of Applied Sciences, Norway
- Anja Møgelvang