

KBL: = Klara Bolander Laksov

RF: = Rachel Forsyth

MF: = Mattias von Feilitzen

CM: = Cormac McGrath

[jingle]

A: Welcome to Högskolepedagogisk spaning, or Higher Education Trendspotting, as we call it today. This is a podcast where we discuss questions regarding higher education, and to our help we have guests who contribute with trendspotting about what is going on in higher education. This is our eighth episode, and our second one in English.

Today's topic deals with artificial intelligence, AI, and more specifically with AI bots' arrival in higher education, and their consequences for teaching and assessment.

I am Klara Bolander Laksov, I'm a professor of higher education, and educational developer at Stockholm University in Sweden, and I will moderate the discussion today.

Let me start by presenting today's guests. We have Rachel Forsyth, works as project manager within higher education development at Lund University. Also, we have Mattias von Feilitzen, educational developer at the Pedagogical Development and Interactive Learning Department at Gothenburg University. And least but not last, Cormac McGrath, educational developer and associate professor with an interest in AI at Stockholm University. You are all very welcome.

We will start with the first trendspotting regarding AI in higher education by Rachel. So, please, Rachel, the floor is yours.

RF: Thank you. I'd like to talk a bit about assessment today, since a lot of the discussion about ChatGPT in higher education has been about assessment and student work. I'd really like to focus on the potential impact of these generative language models on confidence in assessment there, which I think is the main challenge, though I know that people are very preoccupied with ideas of students cheating. Cheating is an artefact of the way that we set up the assignments, so go back and think about that a little bit.

I think there are three key issues here, related to confidence. There is confidence in the process itself, that it's sound and robust. There is confidence of teachers in working with AI tools, and assessing work that may be produced with the help of AI tools. And there is a confidence of students, that their achievements are being appropriately assessed, whether they have used AI tools or not.

So, I think we need to go back a stage to address these issues, to think carefully about concepts of validity, reliability and fairness in assessment. Just recapping those, validity is the extent which a task measures what is the intended learning outcomes. Reliability is the extent to which the hard assessment process leads to the same

outcomes when it's repeated, which is a problematic concept in higher education assessment, where we have complex assignments. And fairness is the extent to which all students have an equal opportunity to demonstrate their achievements and the outcomes.

Now, all of these three concepts are affected by the emergence of these generative predictive text models, like ChatGPT. I don't want to say that these tools are doing things consciously, because they're not. They are just churning out generated texts, which look something like what a human would do. But if you carefully design an assignment to test certain learning outcomes by producing a product such as an essay or a report, then if the student can easily complete the assignment with the shortcut, in inverted commas, a shortcut, if using an AI tool, then reliability and validity really are affected. Because you no longer know what the student can actually do themselves, in relation to that assignment. So, you can't be sure that the learning outcomes are achieved.

In the short term at least, fairness is also affected, because if students have got a differential access to AI tools because of costs or other resources, or if they choose to complete the task in the way that you originally intended, they may be at a disadvantage. So, what many assessment scholars are saying is that we need to look much more closely at process, and much less, at product. And if we start looking at process, so how the student gets to the place where they have achieved the learning outcomes, then the question we can ask is how can these tools be used transparently to support learning, and how can we build them into the assessment process?

Now, students may also have questions about validity, reliability and fairness, if their work is being graded, and feedback is being produced using AI tools. And I think, yes, we need to see that teaching is essentially a relational activity. Teachers want students to do their own work, but students also want teachers to be interested in their work. So, if we focus on that, and start building our joint teacher and student confidence in embedding these tools, maybe this is an opportunity to rethink and refresh approaches to assessment. But it will need all of us to work together, to try and make that happen.

So, I'll stop there, and let the others come in.

KBL: Thank you very much, Rachel. Yes, these are all very interesting questions, and I think that challenge us a lot as university teachers. Mattias, what thoughts do you have about what Rachel raised?

MF: I think that you raise really, really relevant questions and concerns here. And I totally agree that we need to focus even more on process. And I have colleagues of mine, PhD students who are looking into these questions, but not at higher educational level, but in upper secondary schools in Sweden. And they're also talking about these things, going from what he calls a product pedagogy, to more of a process pedagogy. And not looking too much into the products, and aiming for students creating products, and only assessing that. But I also have, well I think it's hard to actually do that in a good way, since we have for many years created educations and settings that actually focus more on products. So, I find it hard within our present educational system, but maybe you have another view on that, Rachel?

KBL: Yes, thank you, Mattias. I think we'll let Cormac comment as well, and then Rachel, you can get back to what Mattias said.

CM: Thank you very much. So, I guess for this conversation today, I've really taken the university teacher perspective, and I think, you know, what Rachel brings up is interesting from that perspective, not least the question of process and outcome. I can see how one can, I can see how to frame process over outcome conceptually, and I see doing it in a K-12 context, whereas as a teacher I have an extended period of time with students, or pupils rather. But just to try in a more, Mattias was saying, and I don't know if that's what you meant, Mattias, I can't help but seeing a lot of obstacles when it comes to giving emphasis to process over outcome in higher education. And I'm not thinking about my own work as an education developer, on the contrary. I'm thinking about people in, you know, computer science, or business studies, that engage with hundreds of students in introductory courses. And I think that we have to be very careful that we don't model practices on K-12, or let's say for me, the worst thing I could do is to model what I say to people who work with students on the type of teaching experiences I have as an education developer, often with very few people. So, it would be interesting to hear more about this process of outcome in particular, or let's say in large class contexts. But I don't know if that's that what you were interested in talking about, Rachel.

RF: I completely agree that this has to be focused on the individual teacher in higher education. I don't have any K-12 experience, so I won't touch on those points. But I think part of the challenge here, and Mattias touched on this, because we've created these settings, but when we say we have created these settings and focus on products, there is an awful lot of myths around assessments. We have to do it this way, because that's how we have always done it. And I think, if you do focus on the individual teacher and their decision-making processes, then it's possible to get people to think about what they can do in their context, whatever it is, big, large groups, traditional focus on certain types of output, and so on. It's only the individual teacher thinking about the assessment that is right for them, which is something you can do in higher education. And it's more difficult in K-12 for the more stretched curriculum.

So, but people need to feel confidence, that's why I keep coming back to these words. People need to feel confident that they're able to produce a valid, reliable and fair assignment that fits their context. And other people can support them with that, but there aren't any imaginary rules about that. So, I agree with you, Mattias, these settings exist, and they have been created, but they can change. And the more we talk about this, the better. But what it comes down to is individual teachers feeling confident in their own decisions about what to do next. And that is quite a challenge for whatever we call it, an academic developer's point of view, pedagogical developer's point of view, that requires quite a lot of input to individual staff. And it's about building their confidence. They know what is best for their students, I really believe that. But they need to feel that we are going to support them in making changes. And that, I think, is a big challenge. But it's not impossible.

KBL: So, can I just add a question then, Rachel? So, how can we work with that confidence?

RF: Well, in the past, I got to pluck my book here, which is called Confidence Assessment in Higher Education, if anyone wants to go and find it. But that text, through a process

called an assessment life cycle, which is a phrase that I coined with some colleagues a few years ago, when we were going through a huge assessment change process in the UK. But basically, there are different stages to assessment management, which are setting and designing in the first place, communicating it to students, actually teaching around it, grading their work. And then reflecting on that and improving it for next time. And it works on the basis that there is no perfect way to manage any assignment. There is a way that works within the context that you're in. And everybody contributes to assessment management, which includes technical administrative staff as well, can do their part to make it work smoothly, and to input their thoughts into it, as well, and students too, of course.

So, by finding a process around, you can make incremental changes, so that it gets improved, year on year. By doing that, people build their confidence. But they need to feel they can do that on an assessment level, and that there aren't like big university-wide rules about things, which usually there aren't. But in people's heads there often are, it's a bit mythical also why we have to do this, we have to have a 3000-word essay. When you look in the regulations, that's almost never true.

So, I think it's getting down to the level of the individual, and supporting them towards, through this process little by little, and building their confidence. And by doing that, we'll manage these challenges that we've got at the moment with AI tools, which can do things that we asked students to do previously.

[jingle]

KBL: Thank you so much, Rachel. I think now it would be really interesting to hear what trendspotting you have, Mattias, so we can add on to this image.

MF: Yes, thanks. So, when I was first invited to join this podcast, I was a bit hesitant. Not because I don't find this interesting, or that I think there are any problems with the pod format, or anything like that. But just because trendspotting, the future of generative AI is really, really hard. And I guess we will feel like we're on this runaway train, trying to figure out what's going on while we're coming up with approaches to handle something that is changing incredibly fast. And things that we talk about today may be more or less irrelevant tomorrow. But then I thought again, and I think there are questions that we need to tackle, and discussions we need to have, that will still remain relevant no matter what changes we see in the future, and how fast the technology advances. And since ChatGPT was introduced just about six months ago, there's been a lot of focus and debate on how to manage these potential issues regarding assessment. And there are some who argue for more traditional written exams, there are some who suggest that take-home exams will work, and if we formulate questions in a better way. And then there are those, including me, who see this as a chance to rethink our assessment formats and designs. But regardless of this stance here, I feel that there is shared longing for tools that can differentiate between AI generated texts and human written texts.

But I think we need some sort of shift here. If we look at these current detectors, AI detectors, they are far too unreliable. We can't trust tools that have mistakenly labelled parts of Macbeth, or the US constitution, or even the Bible, as generated by an AI, of course. And there was a recent study from Chalmers University in Gothenburg, that

surveyed nearly 6,000 students across various Swedish universities, about their use of AI for learning, and their attitudes towards it. And I think there was some interesting findings in this study. Over 60 percent of the students agree that using chatbots in exams is cheating, and yet the majority opposes a blanket ban on generative AI in education. And the study also shows that students are uncertain about what constitutes acceptable use of generative AI in education. And this is where I think we should put our resources. I think we need to produce clear guidelines on what's permissible, and how students should handle generative AI during their studies. And I'm talking, of course, only about AI from an overall educational perspective. There are of course subject-specific AI uses, that's something else. But from an educational perspective.

So, in higher education, I think we must help our students to navigate these technologies. Generative AI will soon be integrated into our most common tools, and therefore it's vital that students learn to use them responsibly. We will soon see GPT in Word, Excel and PowerPoint. Google Bard will be in their Office tools soon. And I fear that students may struggle to understand how to use generative AI appropriately. Especially when it's integrated like this.

So, if we look at other areas where there are problems with students deceiving or cheating during assessment and examination, like questions regarding plagiarism, we have tools like Ouriginal to check students' texts, but that's not our only strategy. We also address these questions through education about academic writing. And this is because we understand that it's also about knowledge and attitudes. So, I believe that we need to have the same approach when it comes to generative AI, and I would like to see universities and faculties and departments digging into this with as much energy as in the discussions about risks of students cheating with AI, and how to prohibit that. And I think that this may be the best way to get students to use generative AI in a responsible way, and also in ways that can actually support their learning.

KBL: Thank you very much, Mattias. You're adding a bit of a different perspective here. Cormac, what comments do you have on what Mattias was raising?

CM: Yes, sure, thanks a lot, and thank you, Mattias. I think that you got me thinking about near and distant futures, and I agree with Mattias that what we're looking at is a near future. We don't know what's coming down the line, as Mattias was saying. A lot of these things might become manifest in the future. And we're being reactive as opposed to maybe being proactive, in the same way we were with respects to COVID. So, we're building strategies to deal with putting out the fire in front of us, rather than thinking about safeguarding for the broader environment.

And I think for, and again, this is maybe less of a trendspotting, so I think my biggest concern, however, would be that there's a, I would make the argument that there is a difference between how professionals and experts use synthesizing, artificial generative intelligence, and how we expect students to learn and engage with material of higher education. To me, there is a big divide that we might be super enthusiastic ourselves, and we see a lot of people would be using this in their professional roles but given that it's not transparent how it puts the puzzle pieces together, I think there is cause for concern there. And maybe we can get back to that in a moment. But those are my initial thoughts.

KBL: Thank you, Cormac. Rachel, over to you.

RF: Yes, thanks, Mattias. I think you've expressed what many people are thinking about and worrying about really clearly now. So, it might be nice to just pursue a little bit how we're going to move from where we are now. I agree with you that relying on detection tools, it may be useful in some cases, but it's certainly not available to us now. And rethinking, while we're thinking about that, we might as well rethink the assignments. But I'd be interested to know how you think, what we can do as a community to support teachers with this, with clear guidelines. Should they be national? Should they be institutional? What should we be doing? We haven't seen any government guides, yes, some countries have published it, but we haven't got any yet. So, I'd be interested in what you think about that.

MF: Yes, thanks, Cormac, and thanks Rachel. I think, I totally agree. If I start with Cormac's ideas, I totally agree with you. This divide is something that we need to actually address as well, I think. And maybe seeing students' use of AI, and generative AI, during their education, during their studies, as also professional, like professional usage, rather than something else. Because that may be, well, have us approaching this in different ways than we do now. For them, they are professional tools, I think. So, maybe that would be more interesting to discuss even further, I guess.

And Rachel, thanks. I think that you're onto something there, because the problem at the moment is we don't know where these guidelines should be formulated. And I think it's more or less the same that always happens, I would say, when something really, when we see really big changes, for example during the COVID pandemic, that all of the higher education institutions started doing everything on their own. And it took some time for us to start sharing with each other and supporting each other. Hopefully, we are in another place now, where we can start supporting each other, earlier at least in the process, and start sharing ideas on guidelines. But who, I don't know who should be in charge of that process. I guess there should be some sort of generic guidelines, or ideas at least, or some kind of support, both for teachers and for students. And then at lower levels we should have more focused guidelines, depending on what subject area it is, and what needs we have. But I guess someone needs to start digging into this, but I'm not sure who should do that.

[jingle]

KBL: Well, who knows, maybe this programme will encourage someone to do it. Well, thank you, Mattias. I think we will continue now, and I'm looking forward to hearing Cormac. What is your trendspotting?

CM: Great, yes, thank you very much. So, I'm happy, so I was introduced, or I introduced myself as an AI interested education developer. I'm also doing a research project, funded by the Wallenberg Foundation, looking at the impacts of artificial intelligence on higher education practices. So, I'm happy to say that to this conversation, I brought in some empirical data. So, together with Klara, and Tessy Cerrato Pargman at Computer Science at Stockholm University, and most importantly, Alexandra Farazouli, we just conducted a research study. And I'll say something very briefly about that. But a quick observation: So, there's a lot of buzz around what chatbots are, and what they do, but I find that the more conversations I'm having at departments,

the more I'm also becoming aware that there is a whole bunch of teachers who were totally blind to this. And that's fascinating. So, it's fascinating, and why it's fascinating I'll get back to in a moment.

So, just to return to this study then. So, in a Turing test-inspired experiment, whereby we had teachers send us home exam questions, and then we started generating some responses to these home exam questions, and then we went to other teachers at the same departments and have them let's say read texts that were previously written by students, and then read texts that were written by chatbots, and they didn't know who wrote what. So, that's kind of the background of the study. And some of the key insights that have come up so far, and these are that across all cohorts we could see a likelihood to pass chatbot written texts at higher than 60 percent. We could see that teachers were very unlikely to suspect chatbot. They had found it difficult to recognize what it is, and there are patterns for recognition, but you just have to wait for the conference paper to come out, to get the good stuff. And we could also find that the current plagiarism detection tools at our university are useless for detecting plagiarism. Because it's not plagiarism as such, if we understand plagiarism as a debate on copying a text. But it is cheating, and it is plagiarism, if we define plagiarism as taking text generated somewhere else, and then pawning it off as your own.

So, then this, yes, had us thinking, and it has me thinking, what does this mean for university teachers and their practices? And I think that one thing I promised to talk about was that I think we are going to see the chatbots will mediate teachers' practices by moving standards. At least if students are allowed to use them in an unregulated way. I mean, chatbots are a form of synthesizing monsters, so they dig through millions of data bits, and then they generate texts based on these prompts. They can have the type of intelligence that humans can't compete with. So, not only does it mimic human intelligence, it's hard to see that in one sense, because it can generate texts at a very fast rate, based on existing texts. And then we can start giving conversations about how sentient and understanding it is. GPT and other chatbots, they are flawless practically in terms of syntax and semantics, that they use correct words, and then we can debate whether or not they use them in the correct manner.

So, I think there is a big, one of the risks, and I'm pitching the conversation today from risks. Maybe I'll come back and, if you like, and talk about the opportunities down the road. But I think that teachers may unconsciously think their expectations on what constitutes a good answer to a home exam, if we suspect chatbot, or even if we don't suspect a chatbot interference. And so, we could also see this in our study that teachers were reading more critically the previous responses by students. Because students make human-like errors. They spell badly, and they do things that are less good than what a chatbot does.

So, I think there is three, let's say, outcomes from this, that maybe have a future-oriented perspective. And as I mentioned previously, I think right now I'm interested in the near and not the distant future. I think we need to examine the policies that are being developed. So, at Stockholm University, for example, we have a central directive that all of the faculties are responsible for acting on this. Which is not really a policy as such, but it's, you know, get going on this. And I think if we don't add to that a component of competency development for those teachers that I mentioned

previously, that have no idea of what's going on, then that puts us in an uncomfortable space.

I think one of my own reflections is that it really calls into question what, you know, the bigger question of knowledge is. And at Stockholm University we are unfortunate enough to have the 7-tier grading system. So, what constitutes a pass at E-level? You know, if I can do that in four seconds on a computer, is this really valuable human knowledge? Obviously it is, but when it's generated by a human. But it starts to provoke these types of questions, what is knowledge and how do we examine it and measure it in the best possible way?

And then another let's say outcome from having done the study is that in some subjects there is a preference for long-format writing. So, deeply engaging with texts written by others, and then starting to form a position of one's own. And if the policies that emerge as a result of this reactive, let's say, reaction to chatbots mean that subjects like, I don't know, sociology, history, philosophy, education, and now I'm just in the, you know, humanities and social sciences, if they start changing their practices away from long-format writing to more snappy, short-format, Q&A-type of responses, what are the implications that will have on higher education, teaching and learning? So, just a few of the things that I've been thinking about, at least.

KBL: Thank you very much, Cormac. Rachel, what are your comments to this?

RF: It has covered so much. I'm touched because of the last point you made there about moving away from long-form writing. I think that's a really interesting topic. So, I was thinking while you were talking earlier that, you know, all the studies have shown that doing something similar to what you did, that there is racial bias in teacher detection, for example, thinking that while written text was unlikely to someone with a minority ethnic origin, in whatever the majority, and wherever it was, and so, yes, universities have added things like Grammarly, which is a form of AI tool as well, to smooth out those things, so, we're trying to help people. So, I guess, do we just assume now that every assignment should have flawless prose, and students can use whatever tool is available? But I am thinking that tools like ChatGPT are not as good at long-form writing as they are at shorter things. So, we might make things worse by doing that. Because quite often, you know, you get outputs that are just several well-written paragraphs put together in an incoherent way. So, I think we're getting a lot of challenges there, we are going to have to test out many things. So, those are a my many thoughts, a few of my many thoughts about what you said. Thank you.

KBL: Thank you, Rachel. Mattias, what do you think?

MF: Yes, I agree with Rachel, there are a lot of things that I would, lots of thoughts rattling in my head at the moment. And the idea that Rachel, the final idea from Cormac, I think it's really, really interesting, something we really need to consider, and have long discussions about. But I'd also like to in some way connect this to Rachel's introduction, her talk, where we talk about process. Because if we are going to continue with these long texts, which I think is really important within certain subject areas, we need, there are even stronger arguments for focusing on the process and following the process. But then we have all these, you know, questions about resources and economics, and budgets, that we need to follow. Somewhere in this we are



supposed to do a good job as teachers, you know. And I think this challenges us as teachers, and it challenges higher education in more ways than just the problem with students cheating on exams. I think that's kind of maybe seeing now, this is maybe the smallest problem that we will have. As I said, we won't know, in two days maybe things are different.

KBL: Yes. Cormac, what do you think? Will we have to make changes like in a few days? Or shall we experiment our way through into the future?

CM: Yes, great questions, obviously. I mean, the study we did was using, so, as you all know, we're talking about generative AI, so something that can create content. And the study we did, we used a version of GPT called the model 3.5. Now, ChatGPT-4 is available at a cost, and it has a much bigger database, or corpus, and can generate much longer responses.

So, I think there's a host of opportunities available for teachers. We can see that contextualizing the material is a way of fooling GPT-3.5. Now, we can see that asking for specific page references, that we, if we work with specific data points, like a section of text, then we can see that it's got difficulty generating responses. So, in one sense, that could and should be part of teachers' competency development, if we think that that aligns well with how we teach in our subjects. Right? But that would enforce us away from, maybe, I'm just saying, the long format. But then there are other alternatives open, I guess. Maybe you don't get three weeks to write a paper. Maybe you get three weeks to do all the reading, and then you have the six or seven-hour open-book exam, where you are given six or seven hours to develop your thoughts. That could also be a practice. Where the process, you own the process, and maybe GPT or other chatbots are part of your process. But then you come to the class, open books, whatever, or not open books, and then you have to show that you can perform. Because there is one underlying element that I think is important, and now I have to use a Swedish concept, just to make sure it's right, but it's this whole notion of *rättssäkerhet*. That we with legal certainty know that the people who are getting a diploma have those skill sets. And in some fields it's really dangerous. Because we wouldn't want medical professionals getting a licence on faulty premises. And then we wouldn't want teachers who didn't know history that are teaching history in our schools, would we? So, I think that we have to be brave enough now to think of the long-term consequences of these tools, as well. While at the same time, seeing them as, you know, there is professional use, and then there is becoming a professional. And if I get one, just a 30 more seconds, I think there is something fascinating in that, the cognitive processes that you need to work with hard, to develop a resilience or grit, or whatever, that you have to return to time and time again. And if we can just ask questions and get answers, that that really constitute a learning in itself. So, I think there are so many interesting things to talk about. But I realise, I'm hogging the word, so I'll stop talking.

KBL: Thank you very much, Cormac, and thank you Mattias and Rachel, as well. I think we have heard many different interesting perspectives, and perhaps, if I'm going to summarise what we have talked about regarding AI bots' arrival in higher education and their consequences for teaching and assessment, it seems to me that we need to focus on assessment, perhaps also, or more, as process, more than only on products

that we have traditionally done, at least since Bologna's entrance into higher education. And to raise confidence and be aware of how we can raise confidence in assessment at many different levels. Also, we need to consider the type of guidelines needed for what or how it is okay to use AI, to enable students to use AI wisely. And finally, AI seems to provoke some bigger questions of what knowledge is, and we need to examine the type of policies that we already have, to be able to move forward, and to perhaps experiment with different ways of approaching assessment, so that we will get that kind of balance between process and product in, at least in assessment. Which means perhaps that we also need to consider this issue of formative as well as summative assessment in new ways than we have done before in the light of AI.

So, with that, I want to thank you, all three of you, Cormac, Mattias and Rachel, for joining this podcast. And I hope to see you again soon.

[jingle]