It's easy to fake engagement with a subject - AI in Higher Education Pedagogy

(Dr. Birgit Hawelka in conversation with Dominik Lukeš)
will say, well, let's just give us a summer at Dartmouth and then a few experts in the field will take the summer and will figure it out. We'll make a long progress. And that was in the 50s.

Now, in the 70s, after many, many ups and downs, one of the AI researchers coined the phrase that's often repeated. His name was Moravec and he coined the phrase calling “easy things are hard and the hard things are easy”. And when it comes to artificial intelligence, all the hard things that are hard for us, like adding numbers and multiplying numbers and reliably storing lots of information, retrieving information, those are hard for us, but easy for computers.

But the easy thing, such as seeing somebody's face like the things we don't even think about, just walking around and opening a window or looking at an apple and knowing it's an apple and just grasping it and looking at it at a word in a dictionary or in a text and understanding all of its subtleties immediately, without thinking, without retrieving anything. Those are things that are easy for us. We don't think about them and the child can do them. Or somebody actually with severe mental deficiencies can, you know somebody who is going to low IQ for whatever reason can do and anybody can do those, but hey, AI can't. So those are the easy things for us or hard. And you all of a sudden this has changed. When the latest iteration of artificial intelligence came around and you said this was very sudden, but in a way it wasn't that sudden. The big change happened in about 2010 or so or 2012 when people started using more efficient computing methods. So, for example, using graphics cards for computing with something called neural networks that had been around since very, very early on, since the 40s, the idea came about artificial neurones came in the 40s.

But all of a sudden we had all the computers and some nice little tricks and lots of data sets. All of a sudden we had, for example, the famous image that data sets came out in about that time. And in a few short years with the image net data set and the GPUs, the graphical processing units, the graphics cards that people are using for games, all of a sudden you could recognise an image, the sort of thing that was thought impossible, right? Just a few years before that, nobody could say, show you a photo of a cat and the computer could ever say there's a cat in it. But now all of a sudden it was getting much better. And it was just that slight change. But there's no intelligence. We just actually started making these easy things for us and hard things for a computer is being easier for computers. And then we just kind of took that revolution.

And since then there was a few more, a few more changes and latest ones came in 2017 with the introduction of something called Transformers and even more efficient training methods and bigger data sets and more sort of thinking about that. And since then we've had this revolution. We had GPT 2 came out and that scared a lot of people. OpenAI, the makers of GPT and ChatGPT and the GPT series even refused to release it to public despite their name open. Because they were afraid it was it would be used for bad things.

And it turned out - and that was about four years ago now - it was so bad and useless that not nothing happened and it was just fine. And then eventually they released GPT 3 and it was it's much less accessible that that's been around for over two years now as well. People could use that. And the big breakthrough came in just a year and a half ago where they figured out that if they just add a bit of training to these models and that training uses human evaluators saying, is this the good kind of question and teach them how to become chat bots essentially, this
big upgrade in chat interface that came out, it was released to the public and the big splash in December of last year. But that made all the difference. And all of a sudden all the things that came together, easy interface, easy to ask and get answers that are real answers to real questions. And the Internet exploded. The people just got started playing around with it. And we're seeing lots of improvements, lots of changes. Since then, other competitors have appeared to OpenAI or competitors, people in that space, other alternatives that we can talk about. That's the journey that we went on. At no point have we added intelligence into the into the mix. Right, you asked: “How intelligent is it?” Essentially we just have bigger data sets and more efficient methods of extracting information from those datasets in a way that we can generate something based on our inputs.

And it goes for images and text as well. So that's a huge revolution in many ways, but it's not an intelligence revolution. It's a revolution in the easy things that are hard for computers all of a sudden have become a lot easier for computers. But no intelligence happened in the sense that we would understand intelligence. Even though the machines can now get a really high score on an IQ test, on a traditional IQ test. But that's only because, again, that was the easy things became easy as opposed to hard. That is the reason. But there is no intelligence module added to this system. It's just still doing the same thing it was doing before, except now we have these bigger data sets and more efficient computing approaches to it.

Birgit Hawelka:
Okay. I see. So it isn't intelligent in a human sense. However, many, especially academics, seem to be quite cautious to chat with this program. Of course many do use it, but others are quite reluctant. For example, the other day I've been talking to a professor who is reluctant to use ChatGPT just because he's afraid that the program learns from each of his qualitative high inputs and thus he might contribute to making the system even more powerful. Does he have a point?

Dominik Lukeš:
No, but a little bit, yes, but mostly no. So essentially, the way these systems work, there are sort of multiple modules stacked on top of each other. And at the bottom of all of these systems sits something called large language model. And that has all the code, intelligence, all the sort of model of the world knowing what comes next. And those models are incredibly hard to train, incredibly expensive, and they need really good data that's reliable, that's huge amounts and reasonably clean data. And for example, we don't know exactly, but it's been estimated that to train the large language will underlying GPT four cost millions and millions or maybe even $100 million and maybe weeks to months of training. So essentially you have a computer churning through all this data over and over again and so intricate ways and which means that to make it give it new knowledge, you cannot do it live. There is no online learning. So essentially, it's not like people imagine humans are. And that's another difference between humans and ChatGPT. Humans learn every time I say something, I'm learning something as well. I'm changing, but ChatGPT doesn't. There's no change whether I ask it. However, there's another layer of this, and it's called reinforcement learning. And that layer has
that's actually human evaluators give it both really highly curated, high-quality examples, and then they rate the responses. That's the third layer.

Now that is also not happening live. So it's not happening as you were chatting in. But ChatGPT is in their terms of service, say we may use some of the data that you use to help us train the model in the future. But it wouldn't be training that sort of underlying large language model, but it might be contributing and we don't know exactly how. So there's a little bit of a justified concern, but not not in that way that somebody may imagine that as you're typing in is going to make that much of a difference. And any one individual is unlikely to actually make that much of a difference.

Birgit Hawelka:
Okay. I see. Whatever we think of ChatGPT, it looks like it's here to stay. And students will use it to support their academic studies. And when we think of academic work in general, we think of mainly three core areas: looking up information, reading and writing. That's what academics usually do. So let's have a closer look to these academic tasks. And perhaps we should start with looking up information. Of course we can ask ChatGPT or other AI any question and get an answer. But how much can and should we trust the information ChatGPT provides?

Dominik Lukeš:
We should trust it 0%. We should actually stop even thinking of this as an information retrieval interface. And it's very tempting to do that because you can have a conversation just like with a human. And very often, more often than not, possibly even it's going to give you correct information. But that's not what you should be using it for because of the famous problem of hallucination. So the way ChatGPT works or these large language models work is: all they're doing is predicting the next bit that comes, and it's not even the next word. It's just something called a token. Which is sort of a part of a word is derived through a process that essentially is compressing the data into something manageable. And it's going every time it adds a token, it predicts the next one and next one the next one. There's a logic to it essentially just saying, okay, this is what might come next and this is generating this. Essentially, it is always say, okay, this might come next, that looks okay. That essentially means it's people say: “Sometimes it hallucinates”. But actually it's hallucinating all the time. All it is doing is hallucinating this next bit, next bit, next bit or whatever, and it always comes with something that makes sense. Since that's plausible, but it has no sense of truth inside, right?

So again, you ask about that intelligence question earlier. One of the features of our experience of human intelligence that we have this sort of pauses, we can say, okay, hold on. Does this make sense or can I break this into steps or can I go look something up somewhere? ChatGPT has no look up mode. It doesn't have a database. It cannot have look into it actually doesn't even have access to its own training data. So there is nothing like that is happening. It's just literally saying, okay, here's a bit of series of tokens, let's generate some more. To you it looks like a sensible thing, like something that makes sense, like a question. And then you get something that looks like an answer. But the ChatGPT under the hood, it just looks like a series of tokens that are all numbers. They're not even letters just for it. It's just numbers.
They have these really complicated vectors assigned to them that indicate the relationship between them and it just keeps spitting them out. And it looks to us like there's some sense going on, but there isn't. There is no logic. It's only that prediction. And it just turns out because of the number of parameters that it's using to make those predictions, which is in this hundreds of billions now. All of a sudden it can do this in multiple languages, can do computer code, it can make these sensible things that so often, probably more often than not, actually are accurate as well as plausible. But we should never trust it. This is an unsolved problem, this hallucination problem. And even if it will often say, for example, I calculated this data by some process, it's only saying this is the sort of thing that might appear here. It's never actually doing it. It never does have access to its internal processes in a way that we might think it does. So that's the important thing. So essentially, if you come into ChatGPT as a replacement for Google, then you're using it wrong.

Birgit Hawelka:
So it's no good advice to look up information with ChatGPT.

Dominik Lukeš:
Well, yes and no. It doesn't replace even Wikipedia, right? Wikipedia or any encyclopaedia. That's not what it's for, even though often it will give you the right answer. You just kind of trust it. You always have to check it. But it's very useful to help you process the information that you find. So very often, sometimes the accuracy of the information, it turns out, isn't as important as your ability to process the information and develop new mental structures. And that's what ChatGPT is so powerful, which is why sometimes it actually makes sense to ask it for something that isn't necessarily factual. So, for example, it makes no sense to ask what year was somebody born? It may give it the right answer or not, but just why? You can google that and get the actual answer very quickly. But for example, you could ask it a question about a structure of a concept. So you can say what are the important concepts in the area of working memory. So for example, I asked that and it's going to give you a list of concepts and then you can go keep using that list to help you develop what I've been calling cognitive scaffolding. To try to help you get better insight into this. So you can then ask it vary that list to subcategories, make it longer, make it shorter, make a table out of this. Those are the essential elements.

And so in a way, for example, if you're a student who's preparing to attend a lecture or preparing to read a new book about a subject, it's actually very well worth it. Asking ChatGPT to give you a list of key concepts, lists of difficult terms. Ask it to elaborate on them in different ways to help you prepare for that, and then use your cognitive scaffold to help you understand this new difficult field. So that is actually a very useful way of using it. But my advice to everybody: always use whatever ChatGPT gives you as a hypothesis to test, right? So, so ChatGPT is going to give you very good answers in many fields, but that's not important whether they're correct or not. What's important to you to help you build those structures so that you can then go and use that experience to help you learn better by confirming whether what ChatGPT told you is correct or incorrect. And that's where I think that this is a useful way of using this.
Birgit Hawelka:
So this would have been this next point. If I understand it right, it can help us in reading text. It can prepare a text to read, it can help us asking the right questions to a text. Did I understand it right?

Dominik Lukeš:
Yes. Yes. So there's a number of things it's very good at. So, for example, you can paste in text and ask: make me comprehension questions, multiple choice comprehension questions about it. And again, if you ask for ten comprehension questions, there's a good chance two of those ten are not actually going to be good ones, as in like it may be badly formulated, the options may be bad, or it might just be a question about something that isn't in the text. But again, as long as you think about this, I'm helping myself understand the text better. So the fact that the question is wrong is as long as you know it might be wrong, it doesn't matter because you can do your reading and confirm: Ah, okay, actually ChatGPT was wrong here. Just hallucinated a question. But doesn't matter. It helped me understand the text. So that's that's a very powerful way of using it.

But you can ask it to look up all the people mentioned in the text and so on and make a table of them, or different terms mentioned in the text. Again, you have to expect inaccuracy in there. So but again, you're building cognitive scaffolding. That's what it is. And so, for example, I asked it to generate a short story in Russian because it's also great for language learning. And it was about a girl who went into an apple orchard and instead by mistake rather found a pear. It was a nice little fun story and I asked it to make some changes to it. But then I said, okay, make me a list of vocabulary items that will help me read this text. And it was really good. All the vocabulary was very good except one of them instead of apple orchard. It had a nut tree orchard in there for some reason. Which wasn't in the text. But for me as a reader, this is again, this is stuff that's hard for computers, easy for me. And of course, I know it's about apples. So even if I'm a non-native speaker of Russian, I can say: Okay, obviously this is not in the text. But all the other things it did for me were very useful for me. So that's I think is important. So as long as you remember never to trust anything 100% but use it as a prompt trigger for your own thinking, that you can then engage with the text and the subject you're studying, then it definitely is a worthwhile tool.

Birgit Hawelka:
I think that's a fantastic way to use it in academic work. However, the point that has been most discussed quite intensively the last month is the question of A.I. being able to produce academic texts. For example, many lecturers are afraid students can make ChatGPT write their essays, which would mean that writing essays would no longer be a proper method for taking exams. Can you imagine or do you reckon ChatGPT being able to write essays in any subject?

Dominik Lukeš:
Yes. It's actually one of the things that people have called this ChatGPT moment is the end of homework. And all of a sudden there is no single subject that using ChatGPT could not a
meaningful contribution to your outcome in it. So, for example, ChatGPT itself cannot write an entire essay beyond 2000 or 3000 words. But again, that's not how you should be using it if you're trying to use it to write something. You can use it to generate an outline and then fill out the outline. So for example, I was testing this one of our exams here at Oxford and I would just do the prompt. And I said, okay, here's the task for the exam. And I said: Make me a list of key points that I would like to make. And then I literally pasted each point to say: elaborate on this just for those five points. And I to be honest, I didn't bother with formatting, so I wasn't write a good essay. But I sent it in and said: You know what, this this isn't very good, but it would pass. So that's the thing, right? So all of a sudden we're kind of getting back to that issue of saying easy things and hard things, right?

So I have looked at my initial testing of this. Before ChatGPT was released I was actually testing GPT 3 for this purpose. And I gave it a task of. Write me something about conversation implicature by Grice - that's a philosopher. And it did. I asked for three different versions just to see because it always changes. And and every three times, every one of those times it got Grice's first name wrong. So I was like, John or George. But it was all right. And so I said: okay, well, who cares? And then it was also a paper that he wrote in the 1950s, and it got that right. That's actually he did write that paper, but it was not the paper where they discussed conversational implicature. So I said, okay, well, obviously this is useless. And then I thought about it for a few minutes. I was actually thinking: No, if a student wrote this, I would pass them because I would think they engage with the concepts. Because who cares about what first name of this not quite obscure, but sort of, you know, not very famous philosopher from Oxford what his first name was. And who cares whether he was publishing in the 50s or 60s. So it just doesn't matter to me, right? I'm interested when the students engage with the subject. And then all of a sudden it turns out that it's actually really easy to fake that engagement with a subject.

So it made me both suspicious of the sort of ChatGPT, but also suspicious of my past engage use of essays. I've been thinking about this even before, but it's actually sometimes quite easy for students to develop a skill that makes it look like they're deeply engaging with something, but they're not there. And we've all experienced when we walked out of that exam and think, Well, I got away with that, right? I talk my way around something I really had no idea about,? And then sometimes we had the opposite. So I really studied hard and I think I understand this, but I just couldn't make it to show in that essay or the exam or the conversation I had with the examiner. So we know that exist, right? We've all experienced it. But we're kind of blindly trusting the fact that when we see students sort of write this fluent text about something that actually indicates some learning. But we now know with ChatGPT, you can do it without any learning, without any knowledge, without any sort of intelligence or deep understanding, just by generating this token, by token, by token. So I would say in a way, the essays were never a very good way of assessing students and now they're even worse. Because we know they can be easily made by this tool.

That doesn't mean necessarily that if you write the essay with all the right intentions that it doesn't tell you that understanding. But it turns out that it can also be written without all that hard work. We're just going to some slick ways of getting your way around the things that you don't want to do. So my answer to that would be, well, yeah, let's stop using essays, maybe.
You can still do them in in sort of a examine individuated conditions where, students are just writing. But were they ever very good?

But the other question, the bigger question is actually: what in 2 or 3 or 5 year time? What will happen to the whole idea of academic writing. At the moment, we have this model of academic writing: Somebody sits down types of letter by letter by letter, But it's a sort of idealised model we have, but it's not really exactly true, right? If you look into sciences and people's publication lists, they're one of 12 or 10 or 50 these days, even or 200 authors. And most of those authors never touched this. To be honest, I have my name on a paper that I didn't write a single word in. And also that can happen. And so we're not always doing this, but we have this model of sort of somebody has to type this word by word. But nowadays that doesn't have to be the case. We can use ChatGPT or these generative AI to help us transform a list of findings into text. And then the human has to change the text. Again, we're always testing hypotheses. We can never trust anything as being accurate. But all of a sudden ... We've the tedium of writing and how many academics just despise writing. And I've done presentations and people have come to me and, and ask: Oh, can you just write my paper for me? Because I was just really hate writing. So successful, published, accomplished academics. Some of them love writing, some of them hate it.

And also, let's look at some of the academic writing that's being produced. Some of it is awful. Some of it is hard to read. Complicated because people just are busy or they're just not very good at it. So all of a sudden, why not use this tool that will make it easier to read for everybody as well as make the job of academics easier? And so I'm imagining in 2 or 3 years time, because I'm getting the questions right now, that will be a much more acceptable way of writing. And I happen to like writing, so I haven't found a good way for me to use ChatGPT for writing. But I know some people who struggle with maybe academic English or English in general, particularly scholars around the world. I've talked to people says, okay, well this is going to save us so much money because we're paying proofreaders, right? We're paying copy editors for this work and all of a sudden ChatGPT can help us produce much better text in English than then we could before. And, and we'll kind of be on a more level footing. So again, I'm hopeful for this as a positive change in the future of academic writing because it's not necessarily a straightforward and easy as people think.

Birgit Hawelka:
Yeah, I'm not quite sure that's going to happen. I hope it will. As a native German speaker, of course, I use A.I. to write English texts because my English sounds sometimes a bit clumsy.

For example, I used ChatGPT, gave it a text and prompted: “Write this text as an academic”. I gave the output to a native English speaker and she laughed at it and told me no one ever would write that way. But maybe also here applies: garbage in, garbage out. Maybe the prompt I chose wasn't good.

Dominik Lukeš:
Yes. That's another question. We're developing this new area of understanding, which is how to formulate the prompts for ChatGPT and these other tools. And for this even a new term has appeared called prompt engineering. And it's not engineering, it's prompt formulating. But I
prefer to think of it as context shaping. You need to set up a good context for the tool to generate more, right? So remember, it's really just as context before and use new things. And so if you formulate the context in a way that as is generating new things is you're going to get better results. Well, that's a useful thing to know. So, for example, a famous mode of prompting ChatGPT is called chain of thought. So if you tell ChatGPT, give it some task. So, for example, explain a theory of relativity or something and say, but think step by step. And just because you ask it to do step by step, all of a sudden it sort of looks like it's revealing some inner thinking, but it isn't. It's still just generating text. But because it's generating this this token bit by bit, it's making richer context. So all of a sudden it's making context. That means that it's more likely to get better results.

So again, the same goes with text, right? So when you're trying to write as an academic, maybe you can sort of stay academic in the sciences, academic in your particular field. You can actually now also give it an example. So you can give it: Here's a paragraph from a journal in my article, and then here is the thing that I wrote. Can you rewrite the thing that I wrote in the style of this paragraph? And again, it's not doing anything under the hood that you would be doing in your head. In your head, you would sort of read the first thing. So here are some features, can I replicate them? None of that, none of that is happening. All that is happening is is generating token by token, by token by token. It's just adding a token after token. Note, those are not words. But because you specify that rich context, the predictions the new tokens generated are more likely to lead in the direction that you want. And and that really is the future of this. This is how we need to start thinking about interacting with these tools.

Birgit Hawelka:
Unfortunately, we are running out of time, but please allow me one last question. Of course, no one, not even you can predict the future. But from your point of view, what steps could lecturers or institutions take to ensure a responsible and effective use of ChatGPT in higher education pedagogy?

Dominik Lukeš:
Well, I'd like to first refute your premise. I think I can predict the future. And my prediction is very simple: The future is uncertain. We live in a very exciting time. And for some people that is a very negative thing to say, for some people that is a positive thing to say. But we also live in a fast changing time and new things are coming. New approaches are coming every week there's something new interesting happening. There's so much learning to happen. And one of the things that makes this interesting is because there's so much intellectual excitement about this. But the thing that makes the difficulty for somebody who isn't intellectually excited about this, this is just complete tidal wave of information and noise and things like that. So I would say: The biggest problem for individual academics, people within these institutions to find a way of learning, developing your own practice. So don't necessarily think of this as something you have to think in terms of how students might use it, but think of it as a tool and how you might be using it.

And I actually would almost say in some fields it's almost irresponsible not to use it. Because imagine: if you can be more productive, why not be more productive? Why not make as a
teacher [use of it]. I mentioned that, I showed you this example of generating texts. We never give students enough examples because they're so hard to come by. ChatGPT or these tools can make examples that are really good. And then you can just quickly, as an expert, tweak them, so make sure that they're actually accurate. And it can generate questions about them. So again, we don't give students enough comprehension questions, enough opportunities to practice. All of a sudden you can make more and better learning experiences for your students. So do that!

And all of a sudden you start learning these issues and again, don't approach it. Many people start their first query and there was my first query as well, trying to catch out ChatGPT. Okay, let's show, let's see how stupid it is, how big a hallucination I can make it. But why? We already know it hallucinates. We should not think of it as something that is accurate. We should think of it as something that can make life easier for us and find the ways in which you can do it. And learn the ways in which it's inaccurate or limited, and then build that into the process, right? Just we've done it with calculators. We've done it with everything, so why not with that? So go through that process. So I think that is the lesson for individuals and it's a confusing moment at the moment. So for example, when using ChatGPT ... but just last week a competitor to open AI, the makers of ChatGPT, called Anthropic have released the second version of their chatbot called Claude. And Claude is very similar to the free version of ChatGPT and it can even let you upload a file. But it has a very interesting feature that's different from ChatGPT. It has a very long context window. So I've been talking about context and I haven't really sort of mentioned the fact that it's limited. So in ChatGPT it's a few thousand, 8000 tokens or 4000 in the free version, which is about 75% in terms of English words, less in German and other languages. Anthropic has 100,000 token context window which could be 50,000 to 75,000 words depending on the language. So all of a sudden you can do much more stuff with it. It has its upsides and downsides.

So people need to start thinking about which tools should I be using. How best to learn the basics of of the prompting design or the context moulding, context shaping that you need to do to get the best results. So those are all things you could start doing. But the biggest danger here ... I hear that all the time and I spent a lot of time interviewing academics, interviewing students about how they use ChatGPT. And I haven't found any difference in sophistication. So I think we always hear that we need to teach students how to use it responsibly. But frankly, I don't believe that we're ready to do that. We don't know how to use it well, just ourselves, right? So so again, let's make sure that we know how to do this right. We can talk about plagiarism and ethics, and that doesn't really change, right? I mean, if you say you wrote something and you didn't, well, that's cheating. And that was cheating before and cheating is cheating now. You know, but the more sort of subtle issues around large language models and this generative AI, those we're not equipped yet to teach students how to do it well and most academics aren't. So we need to sort of develop those.

And that brings me to the question what should institutions do? Institutions should start developing capacity internally. So I'm actually right now sort of proposing this notion of an AI lab within generative AI live within the university. We can test some of these things where people can bring their exams and say, okay, what would that look like if I use ChatGPT on this? People could start thinking about what are good ways of developing materials for
students, What are the good ways that we can use it in office work to developing productivity. But also that would be a place where we can keep track of all these fast developments, right? So because they're not there's no one place I can send you right now to to give you all the information, because this is happening in blogs, on Twitter, preprint publications, all of these lots of YouTube channels. There's so much is happening in this multimodal world of information. And I would not expect any one individual to keep track of it in the way that I've been doing. And so so I think institutions should acknowledge the fact that you cannot give a quick checklist and expect anybody to make good sense of it. This is brand new. This is a big change and we need to sort of support that change and keep track of that fast developing situation.

Birgit Hawelka:
That's fantastic on one hand and a little bit scary on the other, I have to admit. And even though there may be a lot of points to discuss, I'm sorry we have to end up our conversation here at this point. Thank you, Domenic, once again for being our guest. Your insights, I'm really sure, have provided valuable perspective and have greatly contributed to a better understanding of ChatGPT. Thank you so much.

Dominik Lukeš:
Well, thank you for having me. Thank you.