Transcript: In Conversation with Jen Ross and Wayne Holmes – Critical Perspectives on AI in Education and Research

NCRM
NATIONAL CENTRE FOR RESEARCH METHODS

[0:00:00]

Jen Ross: Welcome, Wayne, to this NCRM In Conversation. This particular series is focusing on artificial intelligence in social science research and we're just extremely happy to have you here today and to welcome you and hear some of your thoughts about AI research and other things. Can I ask you to introduce yourself?

Wayne Holmes: Sure. Well, firstly, you know, thanks, it's really good to be here with you today. So yeah, I'm an Associate Professor at University College London and my area of interest basically is artificial intelligence and education, so teaching and learning with AI and teaching and learning about AI, but I take a Critical Studies perspective to both those things. Alongside that I also work for organisations such as UNESCO, the Council of Europe, the Jožef Stefan Institute in Slovenia, International Research Centre for AI, and so on. And it's been really interesting lately with all these new developments and lots of new experts appearing so, yeah, it's been fun.

Jen Ross: Yeah, that does sound like fun, and I think we'll probably get into that a little bit as our conversation unfolds. And my name is Jen Ross and I am Co-Director of the Centre for Research in Digital Education at the University of Edinburgh. And you and I have had some opportunities to exchange some thoughts about the discussion that we're about to have but I think this is going to be interesting for me as well as for those who are hopefully watching and listening later on. So thank you again, Wayne, thanks for coming. So I want to start by asking you how you came to be doing this work around AI in education. Can you talk a bit about that trajectory?

Wayne Holmes: Yeah, sure. So I've never left education since the time I was a child through to now, I've always been in some way or other, but for a long time I have been interested in the use of technology in education and, to be honest, at the beginning I was very much excited by the possibilities, and that's what I did my PhD on, for example. But then over the past decade or so I kind of slipped into looking at artificial intelligence in education, the various ways in which it is being used, and slowly but surely I became more critical of the ways in which it's being used and the kind of promises that are being made around it, so yeah. And I've been in and out of academia working for various organisations since then and seen, you know, how we can move forward. You know, these tools are here, they're having an impact and I'm just interested in helping to ensure that the impact is a positive one.

Jen Ross: Do you still feel optimistic about that possibility?

Wayne Holmes: That's really good, do I feel optimistic? (Laughing) I'm on the fence on the optimism/pessimism thing, I think, at the moment. I think there are so many examples of the way in which these tools are not being used well and the way these tools exist, you know, who owns them, who constructs them, how they're being offered, etc. Are there positive possibilities for the future? I'm guessing there must be some somewhere and I can think of a few but the balance is still not there for me.

Jen Ross: Yeah. And why do you think you became particularly interested in the emergence of AI amongst all the different technological educational possibilities?

Wayne Holmes: Well, as I say, I've always been interested in the use of technology in education more broadly and it just became really clear, as I say, about ten years ago that AI was developing at a pace and was having more and more of an impact, and so, yeah, when I started to get involved I was very excited about the possibilities. You know, for example, an early project I worked on was using AI tools to give students feedback when they were engaging with an online exploratory learning tool for learning fractions, and I can tell you, when I've stood behind a child who was using it then at a certain point I thought to myself, "If I was teaching this child I would say something now" and the computer did give a message that was very similar to what I would've said, that was very exciting and I thought, "This is amazing, what these tools possibly can do." But I think over time it's just became clearer, you know, where the issues are.

Jen Ross: So balancing all of that up, where do you see or how do you think AI has had the most impact or effect on education so far?

Wayne Holmes: Well up until, you know, 18, 20 months ago I'd have said it hasn't had much of an impact. The biggest impact was the emergence of massive multimillion dollar funded companies around the world making massive promises about what these tools can do in education but an incredible lack of evidence to show that they did do that. So don't get me wrong, the AI in education research community have done hundreds, if not thousands, of studies and so, yes, we have seen how these things might be used but the expression I use is that we have no independent evidence at scale for the effectiveness, the safety or the positive impact of these tools in classrooms, and I stand by that. Lots and lots of little studies but mostly those studies have been either undertaken by the researchers who have developed a particular tool so, you know, obviously they have an interest in the tool being shown to be effective, whatever that means, or they've been conducted by companies around the world, and, you know, we need more than that. And one of the things that we definitely need more of is… Most of those studies have been very much focused on the efficacy of the tool so, crudely – and this is not true of all the studies – the student will be asked to do some kind of test, they'll then use the tool and then the student will do the test a second time, and maybe there's a control group alongside. And that may or may not show that there is some improvement through the use of the tool but it doesn't show you the wider context, it doesn't show you, you know, what has been the impact of, for example, the mental health of the student while they've been using that tool, what's been the impact on the relationship between the teacher and the student, what's been the impact on the teacher's empowerment, their professionalism, the student's agency and so on and so on. And I think that's, you know, where we need to go. We need to understand this far more, far better, before these tools become really common. Because at the moment they're still not common, the tools I'm talking about, but of course, you know, over the past 20 months everything's changed and suddenly we've got ChatGPT and the many others. And what's happened with these is, whereas most of the tools education technology have been either brought in by teachers or more often by school management or sometimes by policymakers so it's kind of been very top-down, but with these new tools that have suddenly become available – not new but available – they're from the bottom up, so students are using these tools, teachers are using these tools, and I think that's both fascinating and frightening in almost equal measures, yeah.

[00:08:34]

Jen Ross: Yeah. And I think we'll probably get onto that bottom-up nature of generative AI when we come to talk more about research methods in a minute but I think, yeah, that's a really interesting distinction that you're making been the sort of top-down nature of AI technologies up until recently and then the way those are sort of being now accompanied by a more bottom-up approach that is posing, I suppose, a different set of questions for you in your work.

Wayne Holmes: Yeah, I mean, it's been dramatic. And, you know, I find myself disagreeing with a lot of these new experts but having quite challenging conversations, you know, because… Yeah, because these tools, some can seem so exciting and I come in as the… I'm not a Luddite but I'm kind of portrayed as a Luddite, as the person who's spoiling the party.

Jen Ross: (Laughs) Interesting. Well maybe that relates to my next question as well which is that some of the more interesting discussions about AI have been focusing on matters of ethics and social justice and I suppose bringing those into these conversations you're having broadly probably does throw a spanner into some of the works anyway. What's your take or your view on the way ethics and social justice have been addressed in your research and in the research that you're familiar with around AI?

Wayne Holmes: Yeah, sure. So, you know, everybody who talks about AI talks about ethics, right, it's always in the conversation somewhere, but it's always a very small part and it's always… well, it's not always but almost always it's conceived as being about bias, full-stop. And so, you know, everybody talks about bias and AI and how terrible it is and how we shouldn't allow it to go, and of course that's absolutely true, there are huge problems around bias that's moved into AI, but bias is also used as a way of justifying the use of AI. You know, "We should be using…" the argument goes "…AI in education in assessment because it takes out the bias of the teachers," "We should be using AI to decide whether people convicted of crime should or should not go to prison because it takes out the bias of the Judges," and I find that hugely worrying because it ignores the fact that the tools themselves are, you know, full of biases, both in terms of the data that they are built or trained on but also in terms of the algorithms that are written and the way they are written and the kind of ways in which they target particular outcomes and as a consequence they are very biased." But I think the problem is that ethics in AI is much broader than that and we need to think much more carefully and, you know, ranging from who is in control of this technology. You know, we're often told… well, we're not told specifically but the impression is given that AI somehow acts and operates out there, it just exists in its own terms and it's in control of what it does, and that couldn't be further from the truth. You know, humans are involved at every step of the process, ranging from deciding what they want the AI system to do through the choice of the data to feed into that system, the writing of the algorithms, in a lot of work the identification of the data. And so there's the stories of the so-called 'ghost workers' in developing countries like Kenya who are tasked with looking at huge amounts of data generated by cameras on autonomous cars, for example, but also were used in the early versions of ChatGPT, and these people are paid really, really badly and see some really quite horrific material leading to all sorts of human problems there, but that, again, is not really part of the story when people are choosing to use those tools, that's not known or is forgotten or ignored.

[00:13:10] And I think one of the other problems when we talk about the ethics… As I say, this notion of AI kind of being out there and being independent, but it's not, it's always applied in a certain domain, so it's either applied in autonomous cars or it's applied in medicine or it's applied in climate research or it's applied in education, and each of those domains have long histories where they've been grappling with the issues of ethics, and so when we bring the ethics of the AI and the ethics of the domain together then we have a massive clash and that's rarely thought about, even by some of the leading AI ethics researchers. I think that's a huge problem.

Jen Ross: Yeah. And I suppose equally significant that people working with AI tools and approaches as researchers should have a grasp of that breadth of the implications of AI ethics that you're talking about here. This is not just domain-specific but also really touches on the way that people can or cannot use artificial intelligence tools in their research on whatever topic. So I guess building on that, what kinds of new methodological concerns or ideas do you see arising from AI specifically for researchers?

Wayne Holmes: Yeah, well I think the first thing we need to think about it to make a distinction between AI in research and the natural earth sciences on the one hand and AI in research in the social sciences and humanities on the other hand, because I think it's very different. I try and get my head around what's different and I think it's that the natural and earth science is kind of one step away from direct impact on humans. It does obviously impact on us but it's one step away, one step removed. Whereas research in the social sciences is not, it's directly on people, involves people immediately. And I think, you know, on the natural and the earth sciences, AI has been used quite broadly for some considerable time on a range of things from, you know, climate modelling, optimising crops in agriculture, predicting volcanic eruptions, drug discovery, protein folding, a whole range of different things that I think, you know, there's lots of benefits that have accrued from that. But I think also there are still massive problems and the problems are often either ignored, not understood, swept under the carpet…

So I've got some examples of that, and this one I think people don't get. You know, during the pandemic – so we're talking three or four years ago while it was in the public eye – everyone was saying, "Well, look, this is the moment when AI is going to come to the front, it's going to show its metal," but actually it didn't. So despite there being a huge number of studies using AI both to try and diagnose COVID or predict what was going to happen and develop drugs, it wasn't particularly successful, it really wasn't. So in my talks I reference one particular meta study but there are others. This one was in nature. And that one found that… it identified more than 2,000 published studies and it's phrase was none were of any clinical use, and I think that kind of disconnect between the public appearance, "Oh AI's fantastic to do these amazing things," and the reality I think was huge. But I think there are other issues as well and when we're thinking about, you know, the use of AI to predict climate change and ways in which we might ameliorate that, which is obviously a critical area, it forgets the fact that AI is increasingly one of the biggest culprits in damaging the environment. So, you know, I've got three examples here which, again, I'll refer to, which I think are amazing. AI within the next five years is likely to be using as much energy as the entire country of India, right, the most pulpous country in the world. A typical data centre, as owned by someone like Microsoft today, uses as much water as 2,500 Olympic-sized swimming pools. And the final one, which I think people don't get at all, is that this fantastic tool that everybody loves to use called ChatGPT, well the training of that emitted as much CO_2 as driving a car to the moon and back. Now, these kind of things really don't come to the front.

[00:18:03] One final example I'd like to mention is this thing about the tools appearing to be so good that what's happening is people are becoming over-reliant on them. And so there's lots of evidence from the medical field where they're using AI for detecting anomalies in x-ray scans and other such things and what they found was when the clinicians first started to use them they were very sceptical but then they kind of came around to it and thought, "Yeah, we can use this but we have to be critical of what we're doing," but now it seems that very often they are over-relying upon them and they are themselves becoming de-professionalised as a consequence. So, yes, AI is being used to some amazing effects in the natural and earth sciences but it's not a straightforward thing at all.

Jen Ross: Yeah, and this feels like there's sort of two dimensions to this set of warnings that you're putting forward here. One is around whether these tools can actually do what's been promised for them and the other is whether, even if they can, they should because they may not be resource-efficient enough for us to in good conscience use them as researchers. I mean, is that a sort of fair summary of those kind of key points there?

Wayne Holmes: I think it is. And I think, you know, to be fair to the researchers, there's a lot of work in how do they make their models and their tools more energy-efficient. That work is going on but still, you know, there are stories coming out of places around the world where the local server centre set up by one of these big companies is using so much water, so much energy that the local town doesn't have enough energy, doesn't have enough water for people to live. So, you know, there are attempts in that direction for sure, you know? These tools are not going away so therefore we do need to encourage that kind of work but still I think it's tickling around the edges, it's not dealing with the fundamental problems, I don't think.

Jen Ross: Yeah, okay. Thank you. And I guess having highlighted those really significant sets of concerns or questions that we need to be asking, are there any kind of AI methods that you yourself have found useful or that you're exploring or anything that you think people should definitely stop using immediately?

Wayne Holmes: Yeah, well back in January last year, so just after ChatGPT had been available for a couple of months, I was commissioned to write a report and I thought, "Well I should stop moaning about this stuff and actually try it," and so I paid the $20 a month and I used it as best as I can. And I can write a decent prompt, I don't think that's particularly challenging. Within a short time I was just getting more and more angry and not using it at all; I just found it so unhelpful. But of course since then, you know, lots of people are using these tools and particularly in writing, and, you know, one of the… I don't know whether it's fun or tragic but the story that's comes out recently is the number of papers that are appearing in which the abstract contains some obviously ChatGPT phrases like, you know, "I don't have access to real-time data in the abstract" or "…as of my last knowledge update," you know, those kind of things are appearing. But, you know, apparently if you look at it that's only really a tiny percentage. But I think where I'm more worried are the tools that people are using and for me they're misunderstanding how these tools work and they are being fooled by the appearance of amazingness. And there's no question, right, you use these tools, they do appear amazing, and what the engineers have done to get to that point is amazing, I'm not criticising that at all, but what they don't understand is how the tools work and what the impact how the tools work has on how we might choose to use them.

One example there, and this is something I've been kind of thinking through… It's quite a challenging one to explain, for me to express, not for others. When we use something like a calculator we put in a calculation and we get an output and that output is either right or wrong. Generally speaking, the calculators are now pretty effective, they don't make wrong calculations, but we could press the wrong buttons and therefore get the wrong output. And part of learning how to use a calculator is recognising when the output is so odd that you need to think about "Well actually did I press the right buttons?" so it's that critical engagement we need even with the use of a calculator. And with the early AI tools, you know, pre the large language model like ChatGPT, a similar thing was happening. So one of the classic studies is getting an AI system to recognise handwritten letters and numbers and so by using those kind of tools then that's what's allowed us to… you know, for handwriting recognition, etc. But, again, with that we can agree, you know, we tried to recognise the latter B for Bravo and either the AI outputs a B or it doesn't, so we can see that quite quickly it's either right or its wrong. But the radical thing and the difficult thing for me to get my head around is that these large language models, they just don't work in that way at all. So they don't try to be right or wrong, that's not what they're about, all they're about is predicting the next word, so based on the words we've seen, looking at our training data, what is the most likely next word. That's it. They're really not trying to be correct. (Laughing) But sometimes, because that training data is so large, the output does somehow seem to correlate with reality so it appears correct and at other times it doesn't correlate with reality so it appears incorrect, but it's not correct or incorrect, it's just how it appears. And I think that's a huge problem because people are using these tools and there's a lot of moaning about "Well sometimes they're incorrect" but if only they understood how the tools worked they would recognise that that in itself is a misunderstanding and that actually they need to think, "Well what does prediction process mean on the kind of things I'm generating?" And hopefully to get people to recognise that just because a piece of text or a video or an image, whatever, looks correct, looks real, looks accurate, whatever words we want to use, we must not see it as being any of those things. Now, they might be useful, they be convenient, we might like to use them for our own purposes, that's fine, that's a different thing, but we need to move away from this notion that they're correct or not correct because it's just… Yeah, it's confusing. We're giving it a kind of authority that it really doesn't have, if that makes sense.

[00:26:18]

Jen Ross: Yeah, it does, and I think it brings in all kinds of really important questions about accountability, about the nature of knowledge and all of these things that as social science researchers we grapple with anyway, right, but it sort of adds another dimension to them. Like I've been hearing quite a bit people sort of putting forward the idea that perahps large language models like ChatGPT could be very useful for summarising, you know, large amounts of qualitative data, and I think the point that you're raising here is that however useful or not useful that might seem to be, there is a kind of fundamental question to be answered about what is going on that has produced that either useful or not useful seeming output.

Wayne Holmes: Absolutely. And I think, you know, with the summarising in particular… I mean, there's tools built on top of tools built on top of ChatGPT so it's getting very difficult to understand exactly what's going on but, you know, there's the PDF summarisers on the one hand, there's also the literature review tools on the other hand. Now, we all know that as academics, you know, summarising a text is a pain, it's time-consuming but it's a process that we go through and in going through that process that's where the learning happens; it's not the thing we end up with at the end. And it's the same with a literature review, you know, fabulous to press a button and, voom, there's a literature review, but it's not the literature review that I would've done and it doesn't draw on the kind of things that I would think is important as a researcher. So when I first came across this I was working for UNESCO and there was a conference and the idea was to summarise what people had said in one of the particular session, and ChatGPT had just come out like a couple of weeks earlier so I thought, "This sounds fun, I'll try it". So I took my seven transcripts and I wrote a prompt, you know, "According to this, issues, summarise this speech," and I was amazed. You know, it was early days, I was amazed by what it output. But in this process I accidentally put one of the transcripts in twice and the key is the two outputs were completely different. And I think that's what people have to understand, it's not that when you put a PDF or something into one of these summariser tools it gives you the summary, it doesn't, it gives you a summary, one of hundreds of possible summaries, and that might coincide with what you're interested in but it very likely won't and very likely will some things. So I think, yeah, it can be really hazardous and I would not suggest anybody uses those kind of tools. Maybe for a quick thing just to move on but if you're doing anything of any seriousness then definitely not.

[00:29:31]

Jen Ross: Yeah, that's really helpful and a kind of potentially useful kind of way forward for people in thinking about their own research workflows, right, like where is the moment where this might be something useful, that is okay. That's really great. Thank you. Well I think we should just wrap up with one final question, and I think you've already touched on so many interesting critical questions that we should be asking when adopting AI in research and in research methods. Is there anything that you really want to kind of emphasise or any, you know, particular questions you think we should be asking when adopting AI in our own work?

Wayne Holmes: The key question is should we be adopting AI in our work. You know, it seems inevitable, seems it's out there, it seems we should be using it, it seems that if we don’t use it we’re being very weak, and I think, you know, maybe over time good and effective uses will emerge in the social sciences but I don’t think we’re there yet, I really don’t. I think the kind of tools that are available now are more hazardous than they are beneficial in terms of the quality of the kind of research that we’re trying to engage in. You know, what we do and what social science researchers do, you know, widely, and humanities research as well, we grapple with some quite challenging issues. What these tools are very good at is superficial responses, so if you’re interested in that then fair enough but I don’t believe that most people are. So I think that’s important. I think it’s important to recognise that it’s very easy to be seduced by the apparent quality of what’s output by these systems but we need to be prepared to dig a bit deeper to take a more critical perspective and to think more carefully about, you know, how they work and therefore what they are outputting and to think about when these tools are useful, and I’m always willing to hear when they are. I can’t put my finger on many examples where I think they are genuinely useful but, you know, time will tell, I guess.

I think, finally, the last thing that we have to do, you know, typically we work with students and I think we have an obligation to help our students recognise the challenges of these tools, what they do, what they don’t do, their potential impact on us as researchers, on human rights, on social justice. You know, one of the classics that people don’t recognise is actually when these tools are being used the companies are extracting data from our usage all the time and they’re using that data to improve the tools and therefore to improve their profits. So, you know, it’s been said a long time now that if you’re getting something for free in this world then you are the product and I think this has just been accelerated by the use of these tools. And so I think it’s really important that our students are aware that maybe these tools provide some kind of a shortcut, and I completely understand that in some circumstances a shortcut for some students is essential, I get that, but if the students are genuinely interested in personal development, you know, becoming the best they can become as researchers, then they need to think very, very carefully about these tools. But I would say the same to lots of my colleagues as well who seem to get too excited about the possibilities and don’t think enough about those challenges.

Jen Ross: Well I think this conversation and the points that you’ve raised will certainly be helpful to lots of people as they engage with this series of In Conversation, so I just want to say again, Wayne, thank you so much for making time to do this. It was a really interesting discussion and, yeah, I look forward to seeing more of you work as it continues to challenge and critique the way that AI is being presented in education and elsewhere.

Wayne Holmes: Thanks, Jen, really enjoyed the conversation and, yeah, look forward to continuing the conversation. I think it’s quite important stuff.

[End of transcript]