·VICTVS·

The VICTVS Podcast: Episode 12 — Finding the Balance: AI in Higher Education

Xanthe Mitten:
Hello and welcome to another episode of the VICTVS podcast. Today I'm joined by my co-host, Kat.

Kat Barnett:
Hello!

Xanthe Mitten:
And in this episode we will be discussing AI again, but we'll be looking at its impact on academic integrity particularly in the context of university education. Last month, we spoke to AI expert and TheWhiteBox author Nacho De Gregorio Noblejas, who has a more positive slant on how AI could impact education, particularly in schools. You can see that full conversation with Nacho in episode 10. But today, we are joined by another AI expert and doctor at the University of Leeds, Dr. Richard de Blacquiere-Clarkson and we're going to hopefully understand his view on AI in higher education. Hello Richard.

Richard de Blacquiere-Clarkson:
Hello

Xanthe:
Thank you so much for joining us.

Richard:
Oh, thank you for inviting me. It's a pleasure to be here.

Xanthe:
Kat, I know you've got a lot of questions for Richard. So, do you want to kick us off?

Kat:
Yeah, sure. I was first introduced to Richard through a talk held by or hosted by an International Day for Academic Integrity where he spoke a lot about the difficulty in balancing helping students use AI so that they're well skilled for the workplace, but also maintaining integrity in higher education and making sure that AI isn't overused or impacts academic integrity in any way. It was a really fascinating talk, and that's obviously why we've reached out to Richard to have him on the show today. So yeah, I think it'd be great if you could explain to our listeners first of all, what your kind of roles are at the University of Leeds and what they involve. I know that you work in the AI Hub, so I think that would be a great place to start if you could explain more about that for us please.

Richard:
Yeah absolutely. So, I lead a new initiative at the university lead the AI Hub, which is very much around community and bringing people together, focusing on staff primarily but increasing bringing students in. And the idea is that, essentially, it's about connectivity. It's about building bridges so people can connect together, but also bonding and strengthening the connections between people. So, the idea is to help provide a situation where it's easier to find people with similar interests. You might be working on something then that very often there are other people in the university who are as well and it would be an excellent idea to get together to compare notes, maybe collaborate and so on. One of the challenges we have is that as well as education being very strongly affected by AI at the moment, we're a very large organisation, over 10,000 staff over 45,000

students. So in terms of actually having a kind of really coherent centralised approach to artificial intelligence, it's a real challenge on that sort of scale. So, this is a kind of big aspect of my AI work, and it revolves around trying to provide and help create a context where experimentation and collaboration around artificial intelligence are encouraged because obviously it's quite new. Well, artificial intelligence has a history going back decades, but as we all know, there's been major changes in recent years. It's relatively new in that sense and it's also, as we know, changing very rapidly as well.

So, the approach is essentially what we sometimes call a kind of rhizomatic approach. If you think about fungi in this sort of underground network of nodes which join up together, what you see coming up the top is the sort of the mushrooms effectively, but the real organism occurs beneath the soil and that's very much focused around having a connected network with particular nodes of more sort of concentrated activity. And my approach in that sense is that I'm almost like a gardener. I'm trying to provide the conditions under which that rhizo will grow and will flourish.

Kat:
So in your talk for the international day of academic integrity, you spoke about the kind of battle that academics are facing in trying to balance using AI ethically with also making sure that it, that students aren't using it to cheat. Could you maybe expand a bit more on this and and tell us about the battle that you're facing specifically at university?

Richard:
Yeah, absolutely. So, in that talk I used the phrase “moral harm” which was deliberately slightly provocative because this is often applied to medical contexts which are, you know, we can be talking about literal life or death, but I do think that there's a comparison to be made in that academics and educators more generally are caught between two competing principles, and it's not really possible to endorse one entirely. Nearly everyone has reservations about artificial intelligence as it currently exists. The ways it was developed...there are allegations and some of those allegations have been proven to be correct around intellectual property abuses, around sort of various other kinds of underhand activities that have been carried out by some tech companies. Again, some of this is unproven allegations. There are some, you know, there's some US court documents which provide some quite robust evidence. There are ethical considerations around bias, around representation, around data security and privacy, and so on, which also apply and there's a lot of reservations as we know about the environment and sustainability. There's also concerns that I think we'll come and talk about like deskilling, that over-reliance on artificial intelligence can have negative impacts on students and on learning. And again, there's some evidence to say that this is a genuine risk. and there's also sort of general political concerns. Some educators, not all educators, have got issues around the political economy, the reliance on US companies, the relationship between those companies, the US government and so on. These are individual opinions and I'm not endorsing them as necessarily accurate, but they are opinions that we do come across reasonably frequently. So, in that context, there's a lot of good reasons to be quite skeptical about artificial intelligence and perhaps to kind of limit your engagement or even to decide that you want to take a principled stand and avoid AI entirely.

And I think you can make principled arguments for why you should do that. On the other hand, we have a responsibility to our students to help prepare them for the world as it actually is, and for the future that we believe is coming into play. And all the evidence is this is a world in which AI is already pretty much endemic, various research reports come out estimating that between 92 and 98% of students are currently using AI in their studies. Now of course, these tend to be self-selected surveys. These are students who are interested enough in AI to fill out the survey. So, the real number should be lower, but lower than 98% is probably still very, very high.

Kat:
Yeah, that's a big statistic!

Richard:
Exactly. And we hear similar things from employers. Employers are looking for AI skills because they see it as a really significant skill set for the future as well. So, if we were to go entirely one way and say well, we feel that there are some principal objections to AI so we won't engage with it, then we will not be able to prepare our students for the world as it is unfolding. And I think it would be fair to say we will have let them down and we'll have failed them if we do that. On the other hand, if we're sort of gung-ho, all in, "AI is wonderful",
that it's unproblematic, and we're just going to, you know, really sort of throw ourselves into it, then we're not paying attention to those really important ethical, sustainability, and political facts as well. So, we have to locate ourselves somewhere between those two poles.

But every position is somewhat uncomfortable, as educators, as academics, as people working in this sort of field, we won't all converge on the same spot. Some of us will lean more on the critical side. Some of us will lean more on the kind of, the enthusiastic side. But there is no space at which we can say this is the correct space, we should all be here, and if you adopt this position you can kind of relax and just feel calm and comfortable about it. And I think that's something we need to talk about more, because what we see in reality is a lot of people are struggling with this discomfort because it's not being acknowledged and it's not being validated, and that makes it hard to engage with AI, which is something that we do have to do. We have to have a reasonable degree of understanding. We need to know reasonably well how it works and how to use it, so we can both figure out, well, how are we going to deal with it in our lives? And also, how are we going to prepare our students and our colleagues to deal with it for the future as well?

Xanthe:
Yeah, I guess it's that point of managing AI critically because I read in one of your articles, I think it was in the Guardian about how students use AI or even academic researchers for history related stuff, but how wrong AI is getting all of this historical context and that people can maybe use AI but they have to use it critically because it is wrong. So I guess what you're saying is like, you're teaching your students how to use AI critically and that's what's really important because if we just tell them to ignore it totally then yeah, we're not helping them.

Richard:
Absolutely. And I think this is important. So, the the article I co-authored with Dr. Lorna Waddington, also University of Leeds, does focus on history, but it's part of a sort of a broader point, which is exactly that you can you can apply AI quite successfully in learning, you can apply it quite successfully in research, whether it's as something that helps you gather information that works with text, whether it's something as a kind of thought partner that helps you work through a problem and idea, it might give you really, really good results, and it might give you something that is just very, very wrong - that's achingly, badly mistaken, and that happens sometimes.

It's difficult to put numbers on this, so I'm going to I'm going to pull a figure out the hand and say something like 10% of the time it will be glaringly wrong, and you will probably see it, which then leaves you with something and, depending on the model and depending on the details, we could be talking as high as 30%, of subtly misleading, subtly mistaken errors of material, where there is something that is not accurate or there's something that's taking you in a direction which is less than ideal but it's not glaring either, and that's the real kind of pitfall for everyone is that it's quite easy to look at that and see that it's presented in a way that looks very confident, that looks quite um professional, that looks quite convincing, and not to see it.

Kat:
Yeah. I think that also fits in with the idea that you briefly touched on of these AI systems having biases kind of built into them as well. Like that would also impact how accurate responses are and that's something that students also need to be aware of.

Richard:
Absolutely. And I think it's really important that we're aware that there are multiple types of biases built into the AIs. So one is that they have representational biases. They the biases that are out there in society more broadly are replicated in the AIs because they're trained on essentially vast amounts of literature, vast amounts of the internet. That material is not unbiased and so the AIs will replicate that because that's a sort of thing that naturally happens. It's not a criticism of the technology or the people making it. That is what's happening, but we also find that they have algorithmic biases built into the actual artificial intelligence itself can introduce new bias and can in fact create bias that isn't already present as well as amplify existing biases as well. And you see this in, there's a long history of, for example, using facial recognition software.

So facial recognition software historically is almost always racist and this is an enduring problem. Nobody wants it, you know, people don't want it to be racist. People put a lot of time and energy and effort into trying to avoid this, but it's a recurring issue and that's because of the nature of the biases that are sort of built into it.

Kat:
Wouldn't it have been nice if we could have built AI with like a wonderful utopian vision of what the world could be like, maybe that would have helped. But obviously, as you said, like people's inherent biases go into these things and then it seems to just spiral out of control, I guess.

Richard:
Well, they do and it can, and I don't want to be too negative about it, because you can, in the same way that an individual user can get some really good results out the back of it, you know, you can do a lot of good things with AI and you know, we see it's not hard to find examples of something that's been really helpful and constructive. The difficulty is, it's also not hard to find examples where it's been harmful as well. And we see this with digital technologies more broadly, and we see it, you know in a sense technology enhances... in a sense, technology increases the space of possibilities. It allows you to do new things that you couldn't do before, or it allows you to do it at scale. AI, in many ways has democratised access to expertise, 'expertise' in inverted commas because it's, I wouldn't, it's not exactly the case that AI has expertise, I don't want to anthropomorphise the technology too much, but very often people are turning it to do to do things which some of us could have done already but not all of us. You know, help me find some material on this subject. You know help me improve the quality of a piece of writing that I'm doing. Help me kind of think about this issue. Those of us who had what we might call kind of greater social or political capital, access to money, connections perhaps through family or colleagues, we could do this anyway because we could reach out to people or we could pay people to do it. Whereas now, a lot of people who never had that opportunity can get an AI to do it.

Xanthe:
Okay. So, being devil's advocate here, I've been winding people up in the office all week about the argument of how is this type of AI, where people like the common person can just have access, infinite access to hard answers, which they may have not previously been able to get, how is AI different to like the calculator or when the internet was first introduced? Because people were terrified of them being like, well, it's going to make people stupid because they've just got all this information at their fingertips now. They don't have to do the proper work, or do you think that I'm just chatting absolute rubbish?

Richard:
That's, I think it's a really interesting question because there are some clear parallels but I think there are some key differences as well. So, some of the parallels are that yeah, a powerful technology comes out and people do worry, they do panic about it. It tends to be disruptive but settle down and be a little less disruptive than we thought. So for calculators, the worry was it's like what are you going to do, like how, you know, how are we going to respond to this and it turns out actually, not that much changed, really massive concerns about it in educational circles, around actually teaching mathematics and students learning mathematics, but it wasn't too difficult to adapt at all.

I think the internet you could argue has been more transformative, but we still had a phase... unfortunately, I'm old enough to kind of have been in education as the internet was becoming popularised. My first teaching job at a university, I was the first person to use PowerPoint. I was ceremoniously handed a box of chalk with my first teaching job, and they actually had like a literal green board that you wrote with chalk on. It's quite funny. So it had more impact, but we went from a change where a lot of kind of worry and a lot of suspicion through to now, you know, if you say, well did, are your students using the internet? I'm like, well, presumably? I'm not, whether they're using it really is not a terribly interesting question to me, but I am quite interested in how. And in that sense, I think we're going to go to a similar position where sort of, you know, did a student use AI in this piece of work? I'm like, well, I don't, I guess they probably did. I am interested in how they used it, but whether they did or not ceases to become a particularly important question.

However, the difference, there are some quite noticeable differences. Calculators are not controlled by a small number of corporations in the sense in which artificial intelligence is, and this is a quite distinct situation. The internet has kind of increasingly been kind of absorbed by a relatively small number of companies and popular sites, but it was originally set up as a kind of, quite a sort of open democratising vision of connectivity. AIs of course, you can have access, there are open source AIs, you can host your own, you can build your own, it's not easy to build your own but you know the these are things that can be done. It is possible but most people's experience and access to AI is through a small number of quite large companies who therefore have a lot of influence over how it's developed and how it's used. At the same time, the nature of the technology itself rather than the companies involved behind it is that it is what people often refer to as a black box. Technically, it's a dark gray box, you can peer inside a little bit, but you can't see it very clearly. But, when an AI gives you a response to something that you say, it's not explainable how it got there, and it's not explainable in principle because what we're used to is with technologies like computers and so on, there's a level at which we don't really understand what's going on here. So, I can, you know, you can sort of take apart a computer. You might know kind of well, that's the CPU and that's the RAM, and this is, you know, the memory does this and this is the storage and so on, but then you kind of break it open and what's, you know, what all this inside? And the answer is, I don't know. You know, to me, this is a black box you know, at a certain level, and the same with the software, I'm like, I've got some sense of what's going on, but at a certain level I'm like, what's this? I don't know. But there are people who do, and you can go to those people and they
can explain it to you and we can kind of count on that. But when it comes to generative AI like large language models, there's a certain level at which nobody knows and nobody will ever know. That vector space where they build that probabilistic map of how tokens or words fit together, in principle, we'll never understand it.

Um, and this is really important because it gives us two things. One is that the results are not repeatable.
If you ask, if you keep asking the same thing of the same model you know in sequence, it will give you subtly different answers. Because it's probabilistic. It will tend in certain directions, but they will vary and occasionally it will give you something radically different. And when you put these together, what this means is, is that the

level of trust you can put in these things is exactly zero, in principle. You can never put any trust into anything that a large language model tells you. It might be excellent. It might be garbage. It might be reasonable but subtly misleading in some ways but every time the level of trust...Nothing.

Kat:
And I guess we can trust a calculator, so that is a very crucial difference.

Richard:
Exactly. Yeah. and so you see, there's a lot of analogies and metaphors used and a sort of you know a calculator for words is one that was popularised over the last couple of years, but there is that huge difference because intrinsically in a calculator that's what we rely on is to be absolutely consistent and to apply the same rules the same processes every time. I mean, what would you use a calculator for if you couldn't, you know, that's why we use it is to give us reliable answers.

Kat:
Yeah. almost opposites in that sense.

Xanthe:
Yeah, absolutely. And like I mean the word that comes up all the time in academic integrity and exams is trust, and trust is one of our big things. So yeah, if you're saying that it's 0% trust in AI...

Kat:
That’s not great.

Richard:
No, no. but you can I mean, I'm reasonably critical of this, but I'm also reasonably optimistic in the sense that good things can be done with AI. There's the potential to do a lot of positive things. Um, but this is part of what's necessary to achieve the positive outcomes is to actually have a really good grasp on what's going on fundamentally and how to deal with it, you know. And again, I said there's a pretty good chance depending on the question and the model and everything else, it might give you something that's very good. You just have to check if it's good.

Xanthe:
Off the back of that, are there any models that you prefer or you suggest? Say like if your students are dead set on using AI, or using it as like a research tool or support, is there anything that you say categorically no, that's one to avoid but this one is better, or this one is more sustainable or anything like that?

Richard:
So we have an interesting, so an interesting situation the sustainability is quite difficult to establish, at the University of Leeds, we use Microsoft software, we call ourselves a Microsoft institution and so there's a strong rationale for students to use Microsoft products because we have a contractual agreement that preserves their data privacy and security. Um, we also work with Microsoft on sustainability and they tell us that they use closed loop water systems, so they're not burning through large amounts of water. They use 100% renewable energy in their data centers for Azure. So, they you know, so in this sense we would say actually there's a lot of positives in this kind of relationship. Now, of course, in reality, people use all sorts of stuff, and they don't always use what you ask them to use and there's lots of specialised tools. So, in lots of scenarios, you say, "Well, actually, you know, this product is a great product in general, but for this topic, you might want to use another one,” and one

of the real challenges you have in education is that there's a bit of a wild west scenario because there's so little regulation around AI. It's very difficult to recommend any products with confidence because it's very difficult to know whether they are in fact secure. It's very difficult to establish what the real environmental impacts of different products are, because most of the companies report data in ways that are not consistent. One company will kind of give a headline figure of the total usage, another one will say, well, it per prompt it's this amount, but we're not going to give you the headline figure. And you know, I wouldn't like to say for a minute that this is designed to create a situation where it's hard to compare them, but that's the upshot of the situation. Um, and this is something that needs to kind of evolve over time because a lot of educators struggle with this. It's like, well, I'm not quite clear. I don't feel 100% confident exactly what I should and shouldn't be recommending to students. But it's never, it's likely the case that in order to have a good understanding of AI, you need to experiment and play with a number of different tools and a number of different systems.

Kat:
Moving on to think about the ways that students use AI at the University of Leeds. Richard, could you please give us some examples of the ways in which academics encourage or discourage the use of AI by students?

Richard:
Yeah, absolutely. So, one of the interesting challenges we have is that, what is and is not an appropriate use of AI is very, very contextual and it depends on a number of factors, and the primary one is what is it that the students are learning. So, whether we're talking about what information are they acquiring, what understanding are they building, what skills are they developing or what kind of competencies and abilities are they practicing and working on, that is your starting point and then AI comes in later. Uh I think very often people go, yeah but what, you know, what should I do with AI? What are appropriate uses and not appropriate uses? And the answer is you've always got to step back and say, what are we trying to achieve here? And this is something that we have to reflect on as educators. But it's really important for students to think and ask themselves this question, like, what am I trying to do? What am I trying to get out of this? And then the next step is,and in what way, if any, would AI help me to do this? Because that's the point at which you say well, actually, in this context for this this learning activity using AI in this way is genuinely beneficial but using it in a diff, in another way, is actually undermining me because it's not building me up. I'm not developing as a person if I do this. But that same use in a different context, it could be flipped in the other way around. So for example, one thing that AI can be quite good of is summarising articles, websites, journal articles, book chapters and so on and so forth. And this can be particularly helpful for somebody who for example is dyslexic and that kind of you know long form text is not particularly accessible to them. Now, the extent to which it's actually suitable to use AI to summarise text depends entirely on what you're trying to achieve. If what, if the point of the learning activity is to improve your ability to summarise text, then you're in a difficult situation because you're essentially cognitively offloading that skillset onto the technology. And while it is still possible to use it to help build up your ability to summarise with the AI producing the summary, it's difficult. Whereas, if what you're really learning is how to take those summarised outputs and work with them, it doesn't really matter if you used the AI to produce the summary because that's the raw material that you're working with.

It's a little bit like if you saying to your students, look, I want you to bake me a cake, right? Bake me a cake. And somebody shows up with a cake in a box that's got the name of a supermarket on it and they go, "Here you go. I got you a cake." And you're like, "No, I didn't ask you to bring me a cake. I want you to bake me a cake. I want you to make one." That, that was the skillset. The, you know, the skillset is not delivery of goods. The skillset is turning flour, sugar, eggs if you're not vegan, suitable, you know, flax seed egg replacements, whatever else you want to use.

Kat:
I feel represented as a vegan. Thank you.

Richard:
Yeah. So top tip, chocolate cake, black beans. Yes. And my personal favourite for stupid egg replacements is Dr. Pepper.

Kat:
Oh my god, this has changed my life.

Xanthe:
We've never tried that.

Kat:
Let's do that.

Richard:
This is honestly the least promising cake mix in the world, which turns out great. Carrot cake with a can of Dr. Pepper.

Xanthe:
It's got my name all over it. I'm so excited.

Richard:
You mix it up and you get this disgusting fizzy like slop which looks like you've just made a terrible error. You put it in the oven. Half an hour later. It's a cake.

Xanthe:
A carrot cake?

Richard:
Yeah. It's amazing. It's really good.

Kat:
This has blown my mind.

Xanthe:
This is the best part of the podcast.

Kat:
Right. I'm going to be emailing you after this to get this recipe.

Richard:
This is now a cooking podcast.

Kat:

I do really like that analogy though of the baked cake and the supermarket bought cake. I think that represents things really nicely. How, obviously, there's some circumstances in which buying something or having something ready in advance is useful, but if the whole purpose is to show a skill, to develop a skill, then that's completely lost by using AI.

Richard:
Absolutely. So, if you're, you know, if what you're really interested in is cake decoration, but you need a cake to start with, it doesn't really matter if you bought it. If you made it, great. That's cool, but you didn't, but it's not necessary cause what I'm interested in is the decoration aspect but at the same time if you said to your students right well, I want you to make me a cake you know, the ingredients list everything else, then somebody goes out and they start sort of, you know, sewing seeds to start growing wheat you'd be like well, you've kind of, you've made a mistake in the other direction! in the same way that if a student hands in an essay or a report to you and they go well, I made my own paper, you're like that's really cool but it doesn't change your grade, that's not what, that's not what we're about.

Kat:
Yeah, I guess it's with AI then, it's very context dependent. I guess it’s use has to be taken on a case by case basis really where the students really thinking about what it is that they want to gain from an activity.

Xanthe:
Yeah. And how to use it to make you efficient but not to miss out on any of the skills that you should be using.

Richard:
Absolutely. And that's the risk is that of this sort of so-called cognitive deskilling, because you will be offloading tasks which actually you really should be doing yourself, or offloading them to an excessive degree. Whereas the opportunity is that you can get rid of those kind of extraneous burdens and say, you know, this stuff that kind of you know historically, I was doing but I didn't really, it didn't have to be me. I can offload some of this and that actually frees up time and energy and concentration for the really important bits.

Kat:
I think one example that I can think of of ways in which I've used AI is interview transcriptions, which is very relevant to what we're doing right now. I don't think it's particularly helpful for anyone to sit and manually transcribe something, and that is a really useful purpose of AI. If it can save on jobs like that then that's really great for everybody involved.

Xanthe:
Yeah. Because we manually check it as well after going, but it just speeds up the whole process. So there's still no room for human, not for human error, but there's still no room for AI error because we're still checking it, but we're just not writing it from scratch,

Kat:
Which would take, it does take forever.

Richard:
Oh, yeah. Exactly. And you've hit that really key point. You said you're still checking it at the end because what you're then putting out is something that's got your name on it that says this is from me. I'm, you know,

I'm providing this as an accurate record. So you've checked it, and I think what you probably find is what a lot of us find is that actually, they're pretty good. That generally, you know, chances are you'll find a few little things, but actually they're generally pretty good. And that you, it's a real genuine time saver.

Xanthe:
Yeah. Absolutely. Perfect. So, we've spoken about the impact that we've seen AI have on students critical thinking and reasoning skills. We know that you work as an academic development consultant on curriculum design at the university. Do you think that there's a way that teaching and assessment can be redesigned to preserve or strengthen student autonomy and critical reasoning in an age of AI?

Richard:
Yeah, I think it absolutely can. There are plenty of challenges with how, and a lot of the challenges revolve around the fact that we've seen this quite rapid change, but we're also seeing ongoing change. So, there's one aspect of saying well okay, so how do we respond to the changes which essentially came about with sort of ChatGPT 3.5 and onwards and the sort of the wave of technology that was the sort of the starting point, but we've also got the issue of saying now with the recent developments with agentic AI you know, we're still working on catching up with the last one and there's a real challenge for universities that applies to all educators, but I think universities feel it quite strongly, is that our processes tend to work on a slower time scale than that.

We are capable of moving very quickly, we found that in Covid, it was not fun. We made it work and we're in a somewhat similar situation here. It's not, it's not a crisis on quite the same level, but there's a need for very, very rapid change and a lot of thinking, and there's a discomfort involved in this, because it's not the kind of pace that people naturally feel comfortable with. Because what the adaptation to AI does is it forces us to go back, forces us to go back and rethink what we're doing at quite a fundamental level. And this is good in the sense that it provides an opportunity to reflect on what we're doing and why we're doing it and to make it even better. Take what's good and build on it and where there's room for improvement, we can then work on that. So, this comes back to I was saying before about having to really focus on what are we trying to achieve here? What are we trying to do? And using that as your starting point, and that's just good education. It's the right way to respond to AI, but it was the right thing to be doing already. And more emphasis on the fundamentals of good education is, is a positive thing but the challenge is the scale and the speed involved.

There are a lot of things we can do. So the the foundational work is getting staff and students to have a reasonably good understanding of what AI is and how it works, because if you don't have that, you can't adapt to it. So, at the University of Leeds, we've implemented a mandatory course for all staff and all students which is relatively short. It's a sort of light introduction, but it gives an overview and it gives a starting point and off the back of that, we then have further training kind of workshops and other sorts of activities to help people build up their AI literacy. So, in itself this gives a kind of fundamental baseline understanding. If you engage with this, you'll have the concepts, the building blocks, but beyond that, you then need to start looking at the different aspects of what you do.

There's a lot of focus across the sector on assessment, because that's where people feel it first. I know you'll be kind of feeling this very strongly as well. But for most people, their first encounter in education with AI has been challenges around assessment. So obviously, we've put a lot of effort into this, but we are also thinking about learning and about teaching as well. A good place to start with assessment is that actually for a huge number of people there is a need for redesign. Whether they want to or not, this is an unfortunate truth is there is a need because AIs are quite powerful and increasingly so, and a lot of assessments historically, and I'm talking about the sector and education in quite broad brushstroke terms, would assess product as a proxy for process

because it was a reasonable proxy. It was a fairly good approximation. You know, we're all familiar with university assignments, which essentially you go away and make something and then hand it in, whether it's an essay or a report or whether it's something else entirely. But AI causes a huge challenge for that, because then of course people can go away and use AI in all sorts of different ways and then hand in a product and we're left going well, I don't know to what extent this represents your genuine understanding of this topic.

And this phrase, “your genuine understanding” is really important. A lot of people talk about “is this your work?” and that's an unhelpful phrase, because nothing is one person's work and it never was. Everything is always building upon the work of others. Everything has some aspect of collaboration. Nothing's done in a vacuum. But whether it's a genuine reflection of your understanding really matters because the worst case scenario is somebody goes away, they they lean excessively on AI, they generate something which they don't really get and then they hand it in and the marker, they may well have suspicions of AI but they can't prove it. The marker will then kind of give them feedback, but that feedback isn't addressing them, who they are, where they're at, their trajectory in life. It's aimed at something else. It's aimed at the piece of work that they handed in that they didn't really make a huge contribution towards.

Xanthe:
So, do you see a potential return of more live exams with invigilators where they test knowledge or understanding rather than coursework? Or do you see something maybe like what you do when you do a PhD and you do a presentation of all of your work and you get live like grilled on that to see how much you understand it? Do you see a return of those types of things or something like that implemented at university or even secondary school level, or do you think it's going to be something different entirely?

Richard:
So I think that what we are going to have to see is a hybrid model for most subjects, most programs most of the time. There's been a very natural trend in a lot of universities as a kind of knee-jerk to go well, I guess we're back to exams then because then we can say with, you know if we physically put people in a room and watch them, we can say with confidence the extent to which they're using AI or not. And there is a logic to that but it's, it's problematic in a few ways. So one is that exams in general are not very good assessment methods for the types of learning we're interested in. There are exceptions, sometimes, an exam actually is appropriate, but in general it's not. There's also a challenge because we have legal responsibilities under the Equality Act to provide reasonable adjustments. So, we're not just insisting that everybody uses the same assessment method, because sometimes it's just inappropriate for some people even if it's actually suitable for many others. But on top of that, we have an anticipatory duty under the act as well to not to put the burden onto people with disabilities and additional needs to come to us and tell us. We need to be actually thinking about this in advance about the kinds of, the sorts of likely needs that are going to come up and plan for that in advance. So what, so the situation we should have legally that we need to have is a reasonably high degree of flexibility around our assessments in order to meet the kinds of needs of people that we can anticipate. You know, we know that in essentially every university, and this applies to schools and colleges, there's quite a lot of people who have autism, ADHD, dyspraxia, dyscalculia, dyslexia, and we know this and there are, you know, thousands, there's estimates that around 30% of students have a disability of some kind, and we have to plan for that in advance. If we say there is one blanket monolithic approach to assessment, we are failing in our legal duties. We are letting down a huge number of people and it is a violation of the law. So that, so when I do sometimes hear people say “well, it's back to exams,” the answer is no, it's not.

Xanthe:

Yeah well absolutely it's something that we see every day at VICTVS. I mean, we have so many students and candidates that we have to support through reasonable adjustments and whatever they need so that they can get that life-changing qualification. So yeah absolutely, that flexibility is something that is so important and being able to be flexible, and institutions like University of Leeds. Yeah.

Richard:
However, so I mentioned the idea of having a hybrid because then so blanket exams just isn't going to work – with the caveat there are some courses that will be more exam and observation heavy than others. For example, if you're involved in medicine and healthcare, there may be a strong, quite a strong element of, in order to know whether you can carry out this procedure, we need to watch you carry out the procedure or a simulation of it. But at the same time, what we can't do is just kind of say well, I guess we're moving everything on to unsupervised assessments, coursework portfolio extended pieces of work and so on where essentially, we don't have any robust evidence of how AI is being used because there's then the potential then to end up people go through an entire degree course or a program of study they get a grade at the end, but we don't have enough confidence that that grade is actually accurate or not. So there's a lot can be done to give a mixed and a varied diet. You mentioned sort of PhD style vivas for some people, that would be a really appropriate way of observing them without putting them in a kind of unduly restrictive situation. It also does, it can play to the strength of people who are more inclined verbally than in writing but it also can disadvantage people for whom that's a high pressure situation and it's actually uncomfortable. So again, it comes back to this flexibility.

In the same way, I think there's going to be increasing emphasis on portfolio style and coursework style assessment where people are building up a sort of a body of work over time. And this is something that's being kind of assessed on more of an ongoing basis as well. Now, in that context, of course, there is no proof about the extent to which AI was being used in it. There is no way to disentangle that and say with absolute confidence, I know how you used AI and I know how you didn't use AI. There are things you can do in design to encourage people in certain ways. You can have conversations with your students and help to explain why it's important to do it in a certain way, but you can't have that proof. So, I think what we're going to see or what we, what we're going to have to see, is a bit of a hybrid model where across a program of study there are periodic exams or other types of supervised assessment. And this will have two effects. One will be that you can say with confidence well, in this assessment I know how you used AI because you were, I was watching you. You know, there were people in the room, invigilators or so on, or you know, we were having a live conversation so I could see quite clearly what was going on here, you know, you weren't like looking down, checking your phone and so on. But it also has another effect which is that if a student is using AI on an ongoing basis, they know that it's not too long before the next supervised assessment happens, at which point if they are relying too heavily on AI, they're going to get tripped up. So, this gives motivation to broadly speaking, stay on track with the kind of uses we're looking for.

Xanthe:
Yeah. They have to be on top of it and actually understand what the work they're producing. Otherwise, they'll get caught out.

Richard:
Yeah, exactly. You know, you can get away with it for now, but in 6 weeks or 3 months or next week or whenever it's going to be, you know, then you're going to have a problem.

Xanthe:

And I guess that's where making sure that both the teachers and the students understand how to use AI critically and properly comes into the benefit, because they'll know from the start not to rely on it solidly and then struggle close to the deadline and try to figure out all this stuff that AI has produced for them. So yeah, it's like both of those things like the hybrid testing model but also understanding AI really well, will hopefully result in the perfect use of AI.

Richard:
Hopefully so, and I think it's important to remember that we're all kind of feeling our way through this a little bit and we need to kind of give ourselves a bit of leeway to experiment and try things that might not be perfect. But we need to apply the same charity to students, for there to be a bit of space for them to be able to sort of try some stuff with AI and to be able to give them the feedback and go, "Yeah, it didn't quite work, did it? This has not ruined your future." You know, you're not, you know, you don't have to reset the entire year, this hasn't tanked your degree programme or anything, but in order to succeed and get the results you want you're going to have to now make some changes. And I think it it's really important that we sort of keep that message in mind and say look, you know, this is a an emerging space in terms of evidence and what works, and we need to give all of ourselves a little bit of grace and charity in trying things out.

Xanthe:
Just before we wrap up, Richard, I'm wondering if you could, it's a topic that I've seen you talk about before, if you could give us a little background on the concept of sovereign AI.

Richard:
So sovereign AI is a really interesting concept because what we see is that the majority of the AI models are sort of owned, created and developed by a relatively small number of companies, and these companies are largely US-based but not entirely. They draw on a relatively narrow sort of socio-cultural groups of people, the sort of so-called Silicon Valley mindset. One of the challenges we have is the companies that produce AI, they shape them, they set the guard rails and they fine-tune the AIs and that shapes what it is that it, that the AIs will talk about and what they won't talk about and it also shapes the types of responses they give. Now, in many ways, this is a good thing because we've seen examples of uncensored AI being used for very unpleasant purposes and I'm sure I don't need to go into any of the examples, but you know, they've made the news. So, there's no question that these companies are doing the right thing in thinking responsibly about how do I put some protection in place and limit them? The difficulty is, is that in an academic context this essentially works like censorship because the AI then will support certain types of activities and not others. It will address some topics and not others.
We've seen this with the generation of, to use a historical example, images of Nazi soldiers. So, there was an outcry, this applied to a number of different AIs, but I'm going to mention Gemini as an example because people used it to generate images of Black Nazi soldiers. And there was a criticism saying, "Well, actually, look, this isn't historically accurate." It then transpires it's actually quite difficult to get the images right. A number of different AIs for a while would refuse to generate images of Nazi soldiers. But the difficulty is, is that then who's making those decisions? I'm not suggesting that there's any negative intent, but do they have the right representative groups of people in the room? And the answer appears to be no, partly because people with humanities backgrounds are very, very underrepresented in these groups. And again I don't think there's any kind of, I don't think this is a sort of deliberate ploy or anything like that, but you end up with a situation where one particular group of people from a relatively narrow background, both intellectually but also socio-culturally as well, are making decisions which shape the availability of AI and what it can do to essentially everybody and this isn't a healthy and constructive approach to this.

Realistically from an education perspective, in order to be able to have AIs which will allow us to address the kinds of topics that we want to be able to discuss and the ways we want to be able to discuss and so on, the most promising approach is to look at what's called sovereign AI, which is essentially to make our own and therefore the control over well what what purposes can they be used for educationally is essentially in the hands of educators. Individually, one university, there are very few universities that can afford to
build a really good AI model on their own, but as a consortium, as a collective, the expertise and ability is there. Working with the government would be a natural step as well. And we have seen some examples globally where this has been successful, including in France. The French AI Mistral is a nice example of sovereign AI. So it is, it's not just a proof of concept, an idea, it has been carried out successfully and it is something we could do in the UK if we put our minds and our resources together to it.

Xanthe:
Yeah, I think if we could create something like what France has done, that would be fantastic and probably the best way forward. I think that's about time to wrap up now. This conversation has shown just how complex all the issues that we currently have with AI in education, but also some of the positives as well and the way that it's going. The future is not bleak, it looks exciting and maybe a little bit unknown. And yeah, as AI continues to be more commonplace, we're likely to see more academics like Richard having to make these difficult decisions regarding what role AI should take in higher education, as well as changes to assessment design to help prevent cheating with the technology.

So, I'd love to say a massive thank you to Richard. Thank you so much for coming on the podcast today. We hope you had a good time.

Richard:
Absolutely. It's been a real pleasure. Thank you.

Xanthe:
And thanks to Kat as well for joining me.

Kat:
You're welcome.

Xanthe:
And as usual, please do visit the website if you're interested in listening to more of the podcast episodes or reading our blogs, we keep them up to date every week. And if you've got any questions at all, please leave them in the comments below. You can listen on Spotify. You can also watch on YouTube. And you can visit us at www.victvs.co.uk.