

Welcome to a special episode of the Spotlight on Best Practice podcast, the first in a series that will address concerns around the use of generative artificial intelligence in learning and teaching contexts.

This episode is hosted by Chloe Cauternay, senior academic developer in LTE and leading on academic integrity.

Our special guest is Dr. Robin Crockett who is the academic integrity lead at the University of Northampton and although originally a mathematician, Robin has become something of an expert in academic misconduct and particularly contract cheating and now looking at AI detection.

We are also joined by Dan Clark, assistant director of digital learning, Julie Taylor, deputy director of learning and teaching and Emma Scanlan, senior academic developer in LTE.

Yeah well for those of you who've not met me or can't remember meeting me, I'm Robin Crockett.

I was reader in data analysis in faculty until just shy of four years ago when there was a restructuring and I can't remember where I was based when I last spoke to you, whether it was still in learning and teaching enhancement or whether I'd moved to the registry which is where I'm currently based and report directly to the academic registrar. So and with a brief not so much on AI but mainly on data intensive or cyber intensive as previous academic registrar put it investigations.

So stuff that is basically pulling data apart. And presumably though an interest in artificial intelligence gets caught up in that. Yes well you know I'm a mathematician by background so I actually understand what the AI detectors are trying to do. Yes. Which is kind of how I got involved when I was briefly on the Turnitin customer AI advisory group or a name like that but I was finding them not giving tremendously illuminative answers to the questions I asked. They fairly obviously didn't think I knew anything about you know the quality of text that AIs wrote and how AI detectors worked so yeah I quit. I had better things to do with my time and effectively be dismissed as somebody who couldn't possibly understand what they were trying to do. Okay so obviously last time we met as in when I was in we were looking very much at contract cheating.

Yes. But the shift has definitely moved to worries about generative AI and how to detect it. So I mean I know you've kind of said in an email but would you mind expanding on what you think the future of detection software might be? What direction Turnitin might move in? What risks are? As I rattle through fairly quickly in a moment there is no future to AI detection. As the AIs continue to improve the text they produce becomes more human-like so the differences between AI generated and typical or aggregate human text just get ever smaller.

There's ever more overlap between the sort of statistical distributions of the stuff you can measure and you know the end point will be that the one is indistinguishable from the other to a very very great extent.

So we've got to find alternatives. What kind of warning should we be giving our colleagues then because obviously they we know that Turnitin obviously said it can detect so I have a lot of staff saying to me why don't we sign up to it and I have to explain why we actually opted out. But also I have staff telling me that they've got evidence so they're obviously putting things through other programs are claiming to be able to detect.

Well sometimes an AI just produces very bad text not necessarily in the in the sort of linguistic quality but will just produce sentences that don't make sense or you

know paragraphs that don't make sense when you know read from beginning to end and I'm sure you've picked up on the fake references that GPT 3.5 the first sort of engine under CHAP GPT was infamous for and that's not gone away. You know I've recently seen stuff you know prepared in the last couple of months that still has faked references and is almost certainly AI generated but our regulations do not demand that we identify whether fake information is human or AI or for sake of argument your pet cat. If you as a student submit something with faked or falsified or fabricated information in it is your responsibility we regard it as a serious transgression and that includes references.

You know some people are of the opinion that a falsified reference doesn't really matter but there's the honesty contracting quotes between the student and the grading tutor isn't there.

Definitely I mean I have no hesitation in calling people out for false referencing in a more traditional sense because I'm like what was the point of using something that doesn't exist or saying you've used something you've never used. That's right if a reference doesn't exist then the student cannot use it to support their argument so saying that they do is you know is an integrity failure at several levels. Okay we can all make the odd mistake so you you know every now and then even the best most honest student might fuff up a reference just in a bit of a hurry at the end doing a check excuse me it conflates two references together so you end up with something that doesn't exist. Yeah of course.

But you know where you've got half the references don't exist.

So what are you doing to mitigate that Robin where you are? Well we deal with the students appropriately severely when we catch them doing it dealing with the misconduct when you see it.

Yeah yeah but are you are you doing anything in relation to things like in relation to things like insisting students provide things like annotated bibliography rather than a straightforward reference list or using particular sources that a module tutors identified as being appropriate or anything are you doing anything sort of in the broader assessment strategy or the broader assessment guidance to mitigate? Yeah our basic policy is that a student can use an AI to help them with an assignment unless the assignment brief says no and I'll come back to that point but if they do use it they've got to reference it because it's a source or a resource.

So that leads to sort of kind of two levels of integrity failure one is not referencing it and that's just plagiarism you know like not referencing anything else and the other one is if there are you know faked references falsify information of any other sort then that's a transgression in its own right and we don't need to prove or demonstrate the you know whether it's human or AI in origin.

So we are encouraging our students to learn how to use this stuff some courses are more active about it than others depending on the course teams and the nature of the subject and I don't have enough information to say more about that other than to make the general observation.

So how would you say we should be upskilling staff then to identify AI generated submissions perhaps but also how to feel comfortable in interacting with generative AI or to talk to their students or engage them in the use of it in a positive way? Yeah well I would hope that every member of staff has now at least had a brief play with chat GPT or one of the others even if they're not comfortable in doing it you know it's been out for 18 months.

I don't think that if there was a paradigm shift 18 months ago in a subject area you would expect members of the course team teaching that subject to be unaware of

you know that paradigm shift you know 18 months on.

So staff really need to at least have a play with it and get a feel for what it can do, what it can do well, what it can't do so well. Now I definitely agree because wherever I have the opportunity to interact with staff we're either giving them examples, getting them to have a go. I'm the course director for the PG Cap so I've actually said to them look I don't mind if you use various co-pilot, Gemini, whatever in these ways but obviously the actual output needs to be your own work but if you've done the research and use co-pilot to help find sources that's not an issue for me just tell me that you've done it so that I know.

So I'm trying to encourage that open transparent use of it and the acknowledgement in the hope that that's something they pass on to their students.

Yeah I mean I you know in fairness to our students there a lot of them are going to be graduating into workplaces where they're expected to be able to use AIs and usefully use AIs.

So know how to prompt the AI, check what the AI gives them, re-prompt the AI, make sure you know error check, make sure any references that are provided by the AI are in fact real because you know if you're in the workplace and you use AI for your employer and you get it wrong and something goes out on the net or is otherwise mailed out to people and it's wrong or got faked information in it doesn't do you as the employer any good.

You don't want people who are inept in this, you want people who are you know as up to speed as it's possible to be.

Robin have you encountered any research or anecdotal suggestions of student perceptions of AI and whether they've changed over time because I think there was a I don't want to call it a moral panic but this sort of December 2023 was here or 2022 I can't remember it was a couple years now but there was the huge explosion of or the popularity of chat GPT and I think lots of students would have started to dabble there but I think lots of students are also quite discerning of it and recognize that actually it isn't all it's cracked up to be it does hallucinate.

It's perhaps better to spend the time doing the work than than using an AI engine.

Have you seen any evidence that student perceptions of AI are starting to change? To a point but I think it's still a little early to tell because we're coming to the end yeah I mean you know it's exam time isn't it students will be moving on in June and early July to come back in late September mid-September whatever October so we're getting to the end of the first full academic year and there are students who've learnt how to use it pretty well and think it's great for doing certain things so in a sense using AI as a glorified essay plans or assignment plans but we're still seeing the less engaged less thoughtful students just shoving an essay title or an assignment brief basically into chat GPT and accepting the first text that comes out and in general it's not very good you know I've seen some real howlers and I'm not seeing anything you know that crosses my colleagues paths by any stretch of the imagination but like cheating we're probably only seeing the students trying to cheat not very well well yeah I wondered what you were doing in those circumstances where the piece of work's not very good anyway it's obvious that they've not done huge amounts of research or or alteration to what they've been presented do you treat that as a academic failure or do you treat that as misconduct the most part that would depend on individual tutors and given that sometimes it's not easy to demonstrate on balance of probabilities that AI has been used or misused I suspect a lot of that is just being failed as academic failure I also wanted an academic failure it's an academic failure they're going to get some feedback and some and an opportunity for resubmission anyway I also wonder we've

been doing a lot of work with our colleagues in the school of creative arts and industries recently and they clearly have a lot of assessment that's not written and they are grappling with similar things where they want to use AI purposefully and constructively with students where it's where it's supportive and helpful in in the creative industries but obviously they they still want to protect their the integrity of their their assessments their summative assessments and I wondered if you've done much work with your colleagues that are in in that position rather than the sort of more written assessed work computing I think a bit the same with coding as well actually um personally no but two colleagues um Rob Howe who's our head of learning technology and Jane Mills who's quite got her job title there have been looking at this so that they would be much better placed much better informed than I am on you know the use of AI in in those in more creative assessment yeah you know design rather than writing writing yeah yeah but I can assure you that contract cheating has not gone away there were a year or so ago a bit longer a bit further back than that saying oh this will just kill off SA Mills and contract cheating I never bought into that because Robin is surely those SA Mills are using the AI well yes oh and they they have been for GPT 3 available as SA bots I can imagine that contract cheating SA Mills as they get better at prompting and re-prompting and using more sophisticated AI tools if anything it will become more prolific because it will get cheaper and cheaper for people to be able to generate um you know academic work of a standard that's potentially acceptable if it's not detected um so I can imagine that's actually going to make SA Mills more prevalent rather than less oh you could be right but don't forget that a significant proportion of bargain basement SA Mills in far-flung countries are not interested in whether stuff passes or fails they just want the money yeah yeah you know so if you go back about 10 years a little bit longer perhaps it cost money to contract cheat but you bought in generally to a writer who knew their stuff knew how to write an essay or a technical report or a you know a whatever type of assignment um may or may not have taken any real pride in what they did but they knew what they were doing and you know spent a reasonable amount of time and charged you know corresponding price but now there's stuff that you can get for a couple of tens of pounds or a couple of euro a couple of tens of euros or dollars whatever and it's dire so how do you think there's been any difference since the government actually made SA Mills illegal because I know you just moved your IP address abroad but did it make any difference to outputs or students using them not that I'm aware of well there was no budget or infrastructure to enforce the ban so you know UK-based companies for the most part have as you say just shifted their company registration and and stuff abroad but you can still from abroad use a .co.uk web address or otherwise give the impression that you're based in the UK you know it's not difficult you cover it in union jacks and photos of Tower Bridge basically yeah and we've seen we've seen some student groups being approached by SA Mills not even from an IP address but in WhatsApp groups and things like that as well so they're using lots of other technology to access students and advertise I don't know whether that's still on the increase but it was really picking up just as I was moving from faculty you know so four years ago and somebody from an SA Mill pretending to be a student yeah you get the student group at the beginning of the year when they don't know who each other is so you just had somebody going on hi I'm X I'm a new student on on the same course as you guys can I join you a WhatsApp group and then you know you've just got access to everything that's being talked about yeah and I do suspect that sometimes it's not been so much to sell services to the students but you scrape information when the students exchange notes and uh and assessment briefs all that sort of thing you know you're just in there um but you know not nice but it does go on so we do try and warn our students to be alert about use of social media in all aspects some are anyway because you know they're far more literate than most of us you know on the staff because they've grown up with it but they're you know some are still surprisingly naive at AJ team as they come to us having gone through secondary school with WhatsApp or something like WhatsApp or Instagram you know dot dot dot through basically their secondary school things they still can be very naive that there are nasty people out there on it and

things and we use the the I can't remember the name of the company but they do the international day of action it was against contract cheating and now it's for academic integrity yes but we tend to do that in our institution with the students union and we go into lectures and we have blogs and use student newspaper and have pledges and all sorts of activities so we're trying to be proactive to try and raise awareness of those sort of issues yeah I mean we we've done that for a few years although I'm not sure it happened this year but I've observed that if students in general perceive those activities as being student-led so broadly organized by the students union with no obvious staff involvement seems to have more impact yes as soon as you start to wheel staff out students tend to think oh same old same old they told us this in induction week it was just as boring then you know and that's why we get the students union to do it because exactly as you say it's otherwise it's just another set of rules and instructions so obviously we've ended up talking about the negative sides and the academic misconduct concerns is there any way we can refocus this narrative I mean to start being more constructive or positive yeah I mean you need to get all student-facing tutors to look at learning outcomes basically and get rid of learning outcomes which are just to do with collecting and collating and presenting information because that's what ai is a good act so if you write if you set an assignment which is you know go and read up 10 sources on x and summarize you're just giving marks away basically but what ai still are still not very good at is or still basically can't do but that'll change is they can't do critical insight or critical reflection or critical evaluation they can reproduce other people's you know stuff from their data sets but they can't actually do it themselves so if you set an assignment brief which asks for some critical evaluation or critical insight you know however you phrase it in the subject context but come at it from a non-standard viewpoint then you're not making it any more difficult for students who are engaged but you're making it more difficult for a student who just wants to cut the corner you know and get the ai to do some heavy lifting whether they declare the use of the ai or whether they're actually cheating and not declaring the use of the ai because if the ai hasn't got anything that's coming at the subject from that sort of left field viewpoint it hasn't got any text in its data set it can plunder you know when i when i played with it that's exactly what happened with me i come from quite a small profession in small science space and when i asked the the ai about occupational science which is quite a new and small it couldn't give me it couldn't give me anything meaningful back so i think you're right some of the more left field stuff all the stuff in a discipline that's very contemporary because so if you can really stay on top of the real cutting-edge thinking in your discipline as well as the things that are maybe a bit unique or unusual in a research sense or an authorship sense then you stand more chance of students having to dig for that information rather than it just all being presented to them yeah i mean you know a sort of i've seen you seen your little hand emma come to you yeah something i've used as a sort of blurry example in previous webinars is things to do with the french revolution and the russian revolution but if you as a tutor have habitually asked students to write about the influences of the french revolution on the russian revolution you know what fed forwards all you might need to do well what in the french revolution didn't influence the russian revolution you're still getting your students to read scholarly sources on both but uh you know i i suspect there are few if any sources in the data sets which sort of have that inverted take on the subject now you can only do that for so long because after a while some students will often have their assignments to essay banks the essay banks will get scraped by the ais and you'll start to get something coming through but so you know you'd have to modify things again so emma thanks um well julie kind of touched on what i was thinking as you were talking julie which was around interlinking the professional development of staff with ai around their disciplinary continuing developments obviously we expect all of our lecturers to be very up to date but we're going through a new re-accreditation for a new framework robin and there's an opportunity in there to make it easier for staff to update teaching content based on disciplinary updates because our old process was quite strict around quality in

terms of um module content and things staff especially newer staff were less confident to update content according to disciplinary expertise and particularly assessments questions they know they can't update assessment types but updating essay questions that kind of thing so just thinking perhaps there's an opportunity there as we design and educate our staff around generative ai and academic integrity also around the importance of interlinking their disciplinary update which is sort of part of their research and academic identities with the updating of assessment content or teaching well teaching content and assessment phrasing if you like as well as a way of ensuring one that it's relevant for students and interesting but also that it continues to stay one step ahead of of that kind of constant data scraping and those students who will share information online as well yeah yeah you know and as part of that re-accreditation get everybody to look at learning outcomes so the learning outcomes are you know appropriately focused but also appropriately future-proofed that's an interesting one for us because we've recently just approved a set of collective course learning outcomes and actually I think they are quite future-proof in a lot of ways because they're framed in the 10 domains of things like you know applied skills or communication or digital so they're not framed in the fheq kind of way the leveled learning outcomes are but the but the domains aren't as such and I think that's quite helpful because course teams can adapt them or change them or rewrite them for module level yeah and Emma and I I know we've had discussions in the past and we've not always agreed wholly but but I think we've got a lot of common ground around the critical thinking skills starting from level zero let alone level four but level four you know but how we how we frame that criticality that's that's kind of understandable for course tutors not in this very structured fheq kind of way it's been interesting hasn't it what language we use rather than kind of understanding demonstrate actually using kind of more the language of criticality right from the get-go is one thing we've kind of been thinking about so that's really useful to hear you talk about that in terms of learning outcomes and stripping away learning outcomes that are too directive to information gathering or information summary summarizing and that kind of thing is is really helpful I think yeah and you know empower and encourage or encourage and empower academic staff to you know adapt their assignments from year to year to still meet the underpinning learning outcomes yes we're trying to do that yes that they must read this year's stuff yes absolutely just to preempt you know complaints from students possibly students repeating years or repeating modules um just basically getting the right to make a complaint because something changed and they hadn't been told so if you make it the expectation that students must read stuff and if they're repeating years repeating modules make sure they read the current stuff and not last year's stuff that they become familiar with then you know that that will we're quite flexible with staff about what they validates what they approve in the approval documentations around assessments so we ask them not to be too prescriptive although we want to know they're not over assessing we want to be sure they're not you know doing something that they shouldn't or that would be really detrimental to students at the same time we we do give them quite a lot of freedom to just name an assessment type rather than a title or a or a detailed brief so that they've got a bit flexibility to change what um students are producing in terms of titles but in terms of objects for a portfolio for example they don't have to tell us what the objects specifically are they have to tell us that it's a portfolio um so they could change the objects year on year if if that becomes more contemporary or more appropriate um and that kind of thing but i think there's still a bit of a way to go with our academics around assessment design not just for generative ai but for making sure they're appropriate meeting learning outcomes that they're designed in a way that's supportive to students and all of that kind of thing yeah inclusive as well and to a point that's always going to be the case case yeah you know we all go through being students and then um you know we kind of learn how to teach or tutor or lecture whatever you want to call it and to assess and you know those first things that we design they tend to stick you know we have confidence in what worked yeah and we don't think about actually um making positive change well you know and it

comes quickly enough can be difficult to determine what's a positive change as well you know um i used to teach statistics to environmental scientists almost without exception none of whom wanted to learn how to do statistics to be environmental scientists and there's only so far you can go to make it fun you know if you're dealing with somebody who just doesn't think they will ever need to understand a statistical data set or you know do the statistics you know you're on the back foot from the off yeah and that's just the way of some bits of some courses i remember having the most amazing statistics tutor who considering he was confronted with a bunch of uh therapists rather than you know rather than mathematicians or he was incredibly patient with us but yeah he did make it fun there were there are ways i think well yes but only to a point because you you've got to be precise enough as well but it's also sharing that isn't it i think one of the things we're not so good at at the university hopefully we're getting better as part of this learning design framework we've now got and we've got some transdisciplinary opportunities as core curriculum now much much more um much more evident in our curriculum i think one of the things we've not been so good at is sharing that kind of thing so using your statisticians to teach your therapists or you know or to teach your historians or whatever rather than trying to use your immediate team to do that information you know to do that teaching so i i'm hopeful we'll we'll get better at that over time yeah emma sorry if this i mean i agree with what julie's just been saying and i do hope that that's the case because also in a way that will help tackle the um the siloing of knowledge that can perhaps make ais even easier because as you become more transdisciplinary you have more perspectives that more unusual perspectives and and that in and of itself can be a tool but i was just going to ask about this sticky issue of proof in academic misconduct because it's a question i get asked repeatedly by staff and and nobody seems willing to hang their colors to the line and say what counts as proof and of misconduct well and that that can be basic as well as conduct using ai both actually and we've got we've kind of established norms like you say check the references download the original see how long the document it has been worked on and copy and issues of staff time there but yeah that's kind of well that that i'm afraid is where we're headed it is going to take more time to do some of this um when i first got involved in academic integrity academic misconduct however you want to look at it 15 or so years ago the main problem was copy and paste plagiarism and we had the original turn it in which flagged up runs words that existed you know in in the data set and predated the student's writing you know the similarity check and that's all it was it wasn't a statement that the student plagiarized it was just a statement that somebody else had used these forms of words previously and it did rely on the tutor or the tutor and the integrity officer to look at what the similarities were and whether for example it was just students using subject acceptable jargon maybe not terribly wisely but just running up a sick you know a couple of tens of percent similarity that way or whether as in a few cases that i dealt with you had a very small similarity school but somehow magically somebody else had written the conclusions word for word a couple of years before the student and you know under circumstances where you've got sentences end to end being similar to what's been previously written maybe with the odd changed word or a changing of spelling from american english

to british english sort of thing the chances of that happening by coincidence are just so small you know you think how many different forms of words you could use to say pretty much anything in english so yeah a sentence or a chunk of a sentence coming up chunks of sentences coming up a few times or then whole paragraphs or substantial parts of paragraphs and things we're almost in a world now though where similarity shows that they've read something and actually a complete lack of similarity indicates that misconduct might have occurred well that's that's where i'm going so we all got used to that and then contract cheating started to take off so we started to see low similarity scores and we started to look for other information and you know that to a great extent still exists if you google a few essay mills you will find one that offers a plagiarism free guarantee and you might find that it has sort of the small print on what that means and all the small print

i've seen in that contacts in that contact says similarity of nought to five percent or up to five percent or less than five percent so things that are nought one two three four five percent you know similar quite possibly aligned with an essay mills plagiarism free guarantee and for a lot of submissions is lower than you would expect you know and you can look at that if you've got a whole bunch of similarity scores you know in the 20 plus or minus range and then a few down about three percent and it's not garbage just written by a very weak student who doesn't understand it's actually reasonably written stuff it's probably contract cheated but the same thing can happen with an ai you know an ai is programmed to pull short runs of words from its data set that's aligned with the prompt from the human user and the previous bits of text it's generated in response to that prompt so you might get two or three words from one source five words from as you know one paragraph or one sentence because it's being composited from several so a low similarity score can also indicate possible use of ai and certainly if it's repetitive so if you get several paragraphs in a row all kind of starting in the same way and all looking a bit like introductory paragraphs you know that can indicate ai so low similarity scores bibliography excluded are good indicators of broadly false authorship but there's another catch with ai generated text if you turn the bibliography on you know in the similarity checking you expect the references to be individual highlighted yep so you might get a red one then a green one then a blue one then a yellow one then a pink one just depending on where turn it in or whatever it is seeing seeing the source so if you get a block of five or six or more references all copied from somewhere as a block that can indicate a ghostwriter an essay mill who will have written a zillion essays on subject x and will have a stock of references so they'll just grab a chunk of a reference list from something they wrote last week or last month on the other hand if you see an individual reference that isn't highlighted at all and it's not very current so it's more than a couple of months old sort of thing think why because you that's not what you expect or if you see something where the title isn't really highlighted but the team of authors is the journal is and they're attributed to different sources then you've probably got an ai thinking i've gone three sentences without a reference i need a reference here according to my programming there isn't a real one i need to make one it'll find a team of authors it'll generate a title it'll find a journal title and it knows about volume numbers and issue numbers and page numbers so that's been my experience is you get or you can have a real reference and then a completely different doi so the doi is real and the reference is real but they're not the same paper yes yeah so so where you get a reference which either isn't highlighted at all or a sort of patchwork highlighted check it out and google is pretty up to date so if you paste the title in and don't get a hit will you start to paste the author names in and you don't get a hit sometimes you can paste the journal name in and you don't get a hit but that's relatively rare and okay one reference is one reference like that is probably not enough to proceed on but two is questionable three you're getting into the realms where the student is either too careless needs to be intercepted just for being too careless or they've done something more serious they should not have and as i said don't spend ages trying to find out which ai did it or even that it's ai treat it as falsified information and make sure that falsified information is treated with appropriate seriousness within your misconduct breaks and not restricted to data you know we rephrased ours to make it very clear along the lines of data and information of any sort just a little bit back just to make it absolutely clear that we weren't just looking at numbers thank you that's really good because i think a lot of the hang-ups of the conversations like asked about proof is can we prove it's ai and actually if we can say it doesn't matter if it's a breach of integrity in general then that that puts us back in of using the tools we had available to ourselves anyway um yeah thank you i think that's a really good point because the last two academic misconduct panels i've been on the focus has been very much do you use ai and the students have basically gone i've never heard of ai i don't know how you can accuse me of that and so you kind of then lose it because the focus has gone onto the wrong thing instead of the actual nature of the

misconduct and what we need to do about it to make it good practice yeah the falsifying yeah falsifying of informations a really helpful way of framing it isn't it the misconduct yeah it's it doesn't matter what the the source is it's if it's falsified it's falsified if it's not real yeah yeah i'll come to you in a moment daniel um yeah i initially pushed that change through when we were starting to see some bargain basement essay mill stuff where clearly the ghost writers have got hold of published data sets added some noise so that the numbers were slightly different but they could use effectively the same narrative as had been published so we wanted something that made it clear that it wasn't just numerical data that you know there was also for want of a phrase factual information that can be falsified too we didn't really arguably we didn't really need it but um we thought it would make things clearer to students that would by data we didn't just mean numbers i think this i think this is really helpful and **** i think this is really helpful for the success module in terms of fake news one of the things we've you know talking to students about is that kind of whole the problem of fake news but then the problem of real news being faked as well and and you know so i think if we follow that thread through into the sort of academic realm of the students um work whether it's in qualitative sense or a quantitative sense no matter whether it's data sets or narrative if it's false it's false yeah yeah that's really helpful Daniel slightly tangential question but i shall ask anyone i'm sort of interested in the political economy of educational technology um given that turnitin's business model is is predicated on text matching and if as you say there is no future to ai detection what does the future hold for companies like turnitin what do you think the future holds i don't know i do not have a crystal ball but i think at some point in the relatively near future they're going to have to take stock and decide what their core business is um i don't know if you noticed oh a month a bit more than a month back i put out on the jisk um turnitin thing that i'd actually uncovered a way of beating turnitin by setting tax transparency in microsoft words to a low value i spotted that one yeah it was low enough that it looked normal to human readers but high enough that however turnitin was coded it just skipped through without being checked and yeah uh you know i'm still waiting to hear back from them that they fixed it they're gonna want to recruit you robin you're gonna be you're gonna be then well they're gonna be headhunting you i've just well i don't think i'm popular at the moment because they're my take on ai detectors which isn't anti-turnitin it's just my take on ai you know as a mathematician physicist mathematician whatever who understands the basic theory um but i did decline to sit down with them about um their authorship engine because i've just done so many meetings about that and they seem to write stuff down and then not ever do anything i just um thought i'm not going to go around this loop again but i do have a meeting tomorrow with um glenn thomas i don't know if he's your rep but he's our rep i don't know quite how turnitin divvy up the country anymore we have a new rep now and it's not glenn no i don't know how they divide it somehow yeah well it used to be broadly southern england and then possibly wales and then northern england and scotland or midlands north and scotland because we were in north hampton we were in the northern bit but um yes um but uh the very helpful and experienced lady um moved on not entirely under her own steam didn't she at the end of 2023 yes so yeah but i do have a meeting with glenn sorry they're trying to recruit me well i just think it's interesting the point you make about them them at some point happen to take stock and and work out what their what their their purpose is now it does feel like ai has refocused everybody's attention it's focused the attention of academics but it's also refocused the attention of people like turnitin to think around well

if we can't detect it what is it we're doing now what is it what's our platform about and that's quite interesting it does feel like ai has been a bit of a catalyst for um for a lot of activity in in in higher education yes but also i was just going to say the problem is is we've got so used to having turnitin that people feel afraid that they don't have that corrupt to turn to yes or generally say they are but the the main problem is with ai text detectors is not so much i mean it's easy to say they don't work and that's kind of says everything and

nothing they do work they just don't work well enough for the purposes we need them for and the particular problem which i have found incredibly frustrating is that every ai detector that i'm aware of has optimized its algorithms to minimize the false positive rate so that's detecting human written stuff as ai but um i haven't really got time now have i i had a few slides to show this has everybody got to run away at two i've got a bit i'll try and work out sharing but basically if you optimize the false positive rate minimize the false positive rate you de-optimize or maximize the false negative and that's an ai written text slipping under the radar tagged as human and from my point of view given my role that is at least as important as false positives you know because if i put something through one of these detectors submitted by a student and it's ai generated but the detector says no it isn't that's potentially a student getting away with something they shouldn't be getting away with but let's try can i share okay i can put that up you should see a screen with a code window on the right hand side yep yep yep that's right good right i just need to get this going so i'll just skip all this bit some pictures right there's a distribution in blue of ai text characteristics on the left and a distribution of human in red on the right and i've shaded the colors to show the central bit um in each case um fading out towards the edges and i've just bundled in a few arbitrary points just to show where individual texts lie there's nothing in the vertical scale it's just in the order of entropy increasing along the bottom an entropy is a mathematical coverall term for what the ai text detectors are looking for to distinguish between ai generated on the one hand and human on the other so when it all started out the ais were not that good in terms of producing human-like text and there was essentially no overlap so you know tech ai text detection was doable then then we come a bit more up to speed when we started to get ais that could produce text that was comparable similar to typical you know human-written text and as you can see there's some overlap there there's a couple of blue points lying to the right of a couple of red points lying to the left in the middle there so there's the potential there to misclassify the right to most of blue points as human and the left most red point as ai but you know misclassification was few and far between then we get to something current-ish and the ai has moved and become a you know and become a bit more human-like again how would we actually tell under these conditions whether a piece of text we know nothing about is human or ai well if it was kind of like poor ai we could be fairly confident because we'd be over at the left hand extreme and if it was good human we could be pretty confident because it would be over at the right extreme of human but we still really now need some rules to help us here and this is where these thresholds come in and if we pick an unbiased one it would sit somewhere like that so you'd have a few ais on the human side and a few humans on the ai side but broadly speaking a majority of human is on the human side and a majority of ai is on the ai side but what most detector companies have done is shove the threshold right over towards central ai region because they're trying to minimize the false detection false classification of something human written as ai but that means once there's only in that case one human written thing on the left side the ai side of the threshold there's quite a lot of ai in blue on the human side of the threshold and it's that that is particularly unhelpful and often relayed deceptively in the marketing blurb if you look at most of the marketing blurb you will see they talk about very good very low false positive rates and quietly don't mention false negatives we could set it the other way around so that's just kind of like the mirror image with the false positives but if we go forward a bit more then everything is just so much closer ai is more human like it's moved a bit more to the right again and at the end point you've just got overlapping texts you know in terms of these properties that the ai detectors try to key on so that's basically where i would assert that we are heading so right i'll come out of that now that's a bit of a whiz through um let's see if i can wow right that's out of sharing but we can definitively say there will be no software that is going to identify ai use because it's simply going to be too similar to human writing yes now then i suppose the next question is what's the purpose of identifying as ai anyhow there have been a couple of companies trying to do this for a while but quite a lot that have

jumped on the bandwagon you know they thought there was a market basically um in my opinion and i don't know whether i will get a chance to discuss this tomorrow with um glenn but turn it in would have been better off bundling some sort of tool as part of their authorship suite and not just a general purpose available tool but it is not actually that helpful because you know here i've got you know north hampton i've got charge of who uses authorship and i take responsibility for training them and using it as a statistical tool not as a black and white yes no answer giver we decided against it here so we did a pilot but felt that we were better off training staff to identify academic misconduct um yes um authorship does do some interesting things quickly that it would take an individual tutor or integrity officer um some time to do you know it can do some things in a couple of minutes that it would take a human being a few hours to go in and out of every submission point pulling down every submission to compare you know over a period of time what a student has submitted looking for anomalies and inconsistencies and if you start to cost staff time you know at a typical saving let's say of three hours time it you know the economics don't look that that bad but i mean at the time we did our pilot we found that the economics were balanced between the staff wasn't that's right wasn't it um clary and dan that we felt that actually for the amount of time it was used when people flagged a assessment they had concerns about and the amount of time it took it didn't add up but it might do more so i mean this was a couple of years ago now so it might become more pressing again later on but what was interesting about what you presented there robin is that it kind of this detection and emma's question then is what's the point then of trying to detect is kind of it goes for me it goes it loops back around to what you started with about the falsifying of information and actually it's the falsifying of information that's more problematic than it is the use of ai or the misuse of ai yeah because that's for me where the misconduct is centered and then it's still misconduct anyway yeah whether it's because of ai or because your mum wrote it for you yes now ai detectors basically compare texts to typical or aggregate human on the one hand and typical or aggregate ai on the other and some of them you know have different ai data sets for whether they home in on chat gpt or bard or gemini or claud or any of these others you know some are optimized or have different optimizations but what it's possible for humans to do and what it's possible for humans to do with software is compare to an individual human's outputs i've got software and i have the expertise to take a portfolio of assignments submitted by a single student and compare them for stylistic similarity and differences because that is a whole different question to compare suspected ai against an individual human who will have their own characteristics and quirks as against this sort of amorphous aggregate typical human which is what ai text detectors are now at some point some ai detection services will start to offer the capability that you can upload a portfolio of texts and ask it to identify the different ones or the ai written ones so effectively get an artificial intelligence to do stylometry and you then get a statistical data set you still need to go through it you know that itself is not proof because you've got to factor in that a student might write some short essays some long discursive essays some technical reports some summaries and things all of which necessitate different writing styles so you know stylometry is not a magic bullet but it is a one to you know essentially a one-on-one comparison rather than a one-on-many presumably at some point ais are going to offer the opportunity for students to upload examples of their own writing and then it will write in their style anyhow well yeah that will come and the early ones that do that won't be very good at it but they'll get better very rapidly yeah i think it boils back down to why are we at university anyway what is good practice and can and maybe making that our starting point because otherwise what's the point of any of it well yes you know a student who is you know who's paying their fee and motivated to get a two one because they're nine thousand two hundred and fifty pounds a year they think is buying number two one you know are they really students if that's

their mindset you know and i've met a few with that sort of mindset and they think basically anything goes they don't see the problem because they're paying for a degree they don't see the problem in paying for the work to get them the degree it

it starts to become a problem when that's the medics or the you know or the people who are going into into human-based vocations that um where their knowledge and skills are integral to the delivery of that vocation that's that's for me where there's going to be you know and obviously in medicine and in other health professions you know they've created a way of testing uh students knowledge and skills in a in a very human way because that's what they do anyway day to day but it does worry me that if you know you see this sort of end game of that how that and they can use ai obviously in a very creative and helpful way for medical technology and advancement but at the same time if you start eroding the knowledge and skills early on you know you're not going to you're not going to gain gain it to use it purposefully yeah to use an ai properly in any subject area requires enough understanding of the subject yeah yeah absolutely so that goes back to chloe's question what does it mean to be a student yeah what's the purpose of your learning and that's that's where we've got to home in on learning outcomes and everything else it's developing the student subject expertise developing the love of learning you know developing continue you know the continuing you know purpose of the experience and of the you know of why universities a enjoyable endeavor for students yeah i always use the analogy of a gym if you buy a gym membership you don't get fit if you don't do the classes so you can pay for a university course but if you don't do it you don't do the work you don't have the knowledge you don't have the the ability to do things or the skills that come with doing the course so i'm just conscious of time so are there any sort of last things people wanted to ask to wrap up here yeah i pasted into the chat um about 10 minutes about 15 minutes ago just the link to um a little bit of testing i did in the autumn so before the last couple of revisions of chat gpt and and claud and things which just highlighted the false negative problem and how easy it was to just ask you know whichever ai it was to please write me an essay using long complicated sentences and then the ai detector just doesn't see it as ai that's brilliant thanks robin yeah and i think it's important well i would really like to do with students is to get them to put the simons into things like chat gpt and it analyzes the outputs for me because i've been marking writing for 20 something years when i look at a chat gpt output it's ios task two it's it's a sort of writing to show off your vocab not to show off your knowledge um you know so and you know i'm doing a doctor at the moment and i put things in and amuse myself because i could see it's only gcse level and obviously therefore absolutely useless for me to try to put into my doctor even if i wanted to but do students have that knowledge no of course they don't because they don't have that background so it's helping them to critique those outputs i think is a really important way forward well i mean i'm sure there are academic colleagues who as part of the assessment strategy are getting their students to do some things very similar to that you know so as well as get an ai to write you a thousand word essay on x and critique it is um you write a thousand word essay on it put it through chat gpt and critique what chat gpt does yeah to that because yeah that is educating our students both in the subject area because they've got to know enough about the subject to do that but also how ai's not exactly work in any detailed level but give them an insight to how you know how the ai breaks down the questions and processes things yeah and we mustn't forget as we said earlier that there are positive tips with generative ai and so i think for example copilot will really speed up writing literature reviews for me that's not an issue and i think that's why we want to move away from what dan mentioned moral panics earlier i think there's a tendency to go into a moral panic around generative ai and that's what we need to try and overcome yeah there was something very scary about copilot i saw in the news this morning um about what microsoft are proposing to do that screen shotting yes screen shotting i can't yeah i've got it was on the beeps so now that is really quite scary you know they're microsoft are proposing that copilot can sort of screenshot and screen grab stuff and go through stuff on your computer and and sort of take time stamps of things you know if they force that through i will have to stop using microsoft software for my job you know i simply couldn't afford that to happen with what i do no well with most academic research the last thing you need yes and confidentiality and documents yeah yeah

okay that's all right that's a whole other discussion it is thank you so much robin
i found that very interesting yeah really really helpful really interesting thank
you and i feel i've got thing concrete things to say to colleagues now that when
they come to me with their concern so that's really good hey well thank you so much
for your time robin really appreciate it no thanks robin it's a pleasure yeah great
to meet you and uh sadly or otherwise you now know where to find me you never know
we might be bombarding you with questions soon we'll try not to make you regret it
yeah and if we can be useful let us know yeah of course yes thank you we'll do all
right thanks very much bye