



Science and Technology Committee

Oral evidence: Governance of artificial intelligence (AI), HC 945

Wednesday 29 March 2023

Ordered by the House of Commons to be published on 29 March 2023.

[Watch the meeting](#)

Members present: Greg Clark (Chair); Aaron Bell; Dawn Butler; Tracey Crouch; Katherine Fletcher; Rebecca Long Bailey; Stephen Metcalfe; Carol Monaghan; Graham Stringer.

Questions 263 - 326

Witnesses

I: Daisy Christodoulou, Director of Education, No More Marking; and Professor Rose Luckin, Professor of Learner Centred Design, University College London, and Director, EDUCATE.

II: Dr Matthew Glanville, Head of Assessment Principles and Practice, International Baccalaureate; and Joel Kenyon, Science Teacher and Community Cohesion Lead, Dormers Wells High School.



Examination of witnesses

Witnesses: Daisy Christodoulou and Professor Rose Luckin.

Q263 **Chair:** The Science and Technology Committee continues its inquiry into the governance of artificial intelligence, looking this morning at the applications of AI in the field of education—specifically, schools.

We are very pleased to introduce our first panel of witnesses. Professor Rose Luckin is professor of learner centred design at University College London. Professor Luckin's research involves the design and evaluation of educational technology that uses AI as a tool to support teaching and learning, and the implications of AI's growth for education policy. Thank you for joining us today.

With her in person is Daisy Christodoulou, who is director of education at No More Marking, a provider of comparative judgment essay assessment software. Before joining No More Marking, she was head of assessment at the Ark Schools academy group. Prior to that, she was a secondary English teacher in London. Thank you very much for joining us today.

The interest in AI and its governance is very intense at the moment. Just this morning, the Government are about to publish a White Paper setting out their proposed approach to the governance of AI. The White Paper invites responses before it is confirmed. The Committee will be feeding into that as well.

May I ask both our witnesses to comment on the headline indication from the White Paper, which is that, rather than give responsibility for AI governance to a single new AI regulator, the Government intend to empower existing regulators? Obviously, there are several that exist in the education sector. Do you have any initial reaction to that, Professor Luckin?

Professor Luckin: It is an interesting way forward. I understand the rationale behind empowering existing authorities, but I worry that existing authorities do not have the in-depth knowledge about the technology really to be able to do a good job. I do not mean that in any insulting way to them whatsoever, but the technology is moving at pace and is increasingly complex. Even the people who are developing it do not always understand the implications of exactly what it does.

I hope that, in parallel with empowering those existing bodies, a tremendous effort will be put into making sure that they have the skills, knowledge and understanding of the technology to be able to do their jobs with force.

We have to recognise—and I am sure that all of you do recognise—that AI is a technology that is not going away. It will only become more prevalent, be used by more people and have even greater impact, so it is super important that we make sure that people understand enough about it to be able to do that regulatory work.



Q264 **Chair:** Indeed. The regulators in the school setting include Ofsted and Ofqual. Are there others that you think might be in scope for this in schools, as part of education?

Professor Luckin: This is a bigger question. I do not think that we have put enough attention into educating people about AI. It is therefore really hard for me to point to other organisations that I feel have the knowledge both about education and about AI. It is a bigger issue. I am sorry if that is not a very satisfactory answer. It is just that a key issue for me is the lack of attention that we have given to educating people about AI, beyond doing a great job of investing in STEM, particularly in universities, to make sure that we have the technical skills for AI—but that is not the same thing.

Q265 **Chair:** The purpose of the inquiry is to make some recommendations.

I put the same question to Ms Christodoulou. Do you have any initial thoughts? I do not think that it has been formally published yet, but we know about this intended approach from what is in the press. Does it sound to you that this is the right approach?

Daisy Christodoulou: I agree with Rose that one of the challenges with the new generative AI tools, in particular, is that they are very complex. She is absolutely right to say that sometimes their own developers do not really know exactly what they can and cannot do. There is a huge level of complexity, so there has to be a concern about whether existing regulators will have the expertise.

Having said that, if we are looking within education specifically, I know that a few years ago Ofqual looked at the potential for AI to mark assessments. I know that it is on its radar. The immediate implication of a lot of the generative AI—the large language models—is probably going to involve assessment, so, if you want to move quickly, it probably makes sense to empower the existing regulators to look at it.

As well as being concerned about the complexity, I am concerned that a lot of organisations are moving very slowly, whereas the speed of the adoption of the technology in society is much quicker.

Professor Luckin: In the tech companies.

Daisy Christodoulou: There are organisations that will end up setting up a working group to think about it in a year's time. By that time, you could have almost every student using it.

Q266 **Chair:** So pace is important.

Daisy Christodoulou: Yes. Speed matters here.

Chair: Good. Given that the White Paper is being launched today, I will take a couple of questions on it before we go into some of your experience of AI in education.



Q267 **Stephen Metcalfe:** Professor Luckin, you said something about people's lack of knowledge and understanding of AI. Do you have any suggestions on how we might improve that generally within the wider public? I think that it is important that they understand what AI is, rather than just take their reference frame from what they might have seen on television or film.

Professor Luckin: Absolutely. We should not underestimate the watershed moment that this is. It is not that ChatGPT-4, DALL-E or any of these generative AI tools are cutting-edge technology—they have been around for several years. The huge difference in what has happened is that it is the first time in the history of the human race that people at scale have had access to an AI tool that they can use. It is very usable. It is there. It is just like when Google first appeared as a search engine, which I am old enough to remember. It was so different from anything else that we had ever had. People are using AI, and they know that they are using AI. They have been used by AI for many years.

We must not underestimate the significance of what is happening, on that scale. Therefore, we have to recognise the scale of the need for education. We must look in schools, in colleges, in universities and in workplace training, and think about what people really need to understand about AI in order to be able to use it effectively—and to stay safe. *[Interruption.]* I am sorry; I have an abnormally high heart rate. I am not sure how useful that is. *[Laughter.]*

Chair: I hope that we have not caused that so far.

Professor Luckin: I am terribly sorry about this.

Stephen Metcalfe: I only asked a small question!

Professor Luckin: Perhaps it is because I feel very strongly that we need to look at education and training across the board. I love programming and technology, but not everybody does, so we have to be careful not to approach it by expecting everybody to want to understand the intricacies of the technology, how it works and what it does—to want to understand the maths behind neural networks, to learn how to program in Python or whatever it might be. That is lovely stuff for those who like it, but we need to find a way to help everybody to want to understand what this can do in very general terms—what it means to be artificially intelligent. I suggest that a way of doing that is to think very carefully about what it means to be intelligent as a human, and then to understand the differences.

Q268 **Stephen Metcalfe:** Absolutely. I have a final tiny question. How do we teach, using our own intelligence, a degree of scepticism? As an experiment, I asked ChatGPT to write a 300-word biography about me. It did so, but it was inaccurate. It was me, but it was inaccurate in all the detail. No one else would know that, necessarily, so it almost needs to come with a health warning.



Professor Luckin: It does. I have to laugh. I asked it about AI in education and whom one should talk to about that. I felt very flattered when I was No. 1 on the list, but I do not think that that is accurate. I had the same experience.

Yes, a healthy scepticism is important. That comes from understanding that a tool like ChatGPT does not understand a word that it produces. It has no conception of how those words apply in the real world. Once you grasp that, you can start to become much more sceptical. Once you learn how to write the right sorts of prompts—the prompt engineering, as we might call it—you soon find out exactly what it cannot do.

It is incredibly useful. I use it. It is ever so useful, but you need to recognise that this is information collation. It is not knowledge. It is not understanding. It is collating vast amounts of information, and often incorrectly.

Stephen Metcalfe: Absolutely.

Professor Luckin: It will get more accurate, though. I will drop off the list then.

Q269 **Stephen Metcalfe:** It will. As someone once said, there is nothing artificial and there is nothing intelligent about AI.

Professor Luckin: No.

Chair: Dawn Butler has a question on this theme.

Q270 **Dawn Butler:** Daisy, you said, “If we need to do something quickly, this is the way to do it.” Is it sustainable, and is it the best way? Are we going to get the best results? Are these organisations equipped to be able to educate people, as Rose was saying, or to protect people with regard to AI?

Daisy Christodoulou: That is a really good point. I have to say that the speed at which it is moving is quite disconcerting. It has been disconcerting for us as an organisation.

To go back to the previous point about education, one of the challenges is, how do you educate people when it is moving so rapidly and when there are genuine debates between people within AI about exactly where it will end up?

I am not going to make huge predictions, but the important thing to note is that people have quite strong disagreements about where it is going. Some people will say, “It is only going to get better. Yes, it is making mistakes now, but it is going to be pretty flawless within a year or two.” Other people are saying, “No, that is not the case. It does not have true understanding. Therefore, some of the things that it is producing are a bit like magic tricks. It is not going to keep developing at pace.”



HOUSE OF COMMONS

I guess that the reality is that we do not know. Educating people and regulating are hard with that level of uncertainty.

With some things, perhaps, you do not want to move quickly, because of the uncertainty. In some cases, there is an argument for slowing down. The places where regulators have to move quickly are those areas within education where it could have an immediate impact, and is probably already starting to have an impact—in particular, if students are starting to use it to write and hand in work. That is the most pressing immediate need. You would say that regulators, schools and universities need to be thinking about that and about their assessment models.

I do not always think that there are easy answers to some of those questions. They are not straightforward. That is what I think needs to be high up on institutions' radar.

As I said, it worries me when I see institutions thinking, "We will put together a working group about this in six months." No, it is happening now. You need to be thinking about this now. There may not be easy answers. It does pose challenges, but in those areas that are immediately impacted by AI we have to think about it quickly.

Q271 **Dawn Butler:** What is the one thing that we must do in this area to ensure safeguarding?

Daisy Christodoulou: That is a very good question. What is the one thing that we must do?

People have discussions, and schools have a lot of autonomy in what they do. I am thinking about this specifically in the school and university sector. Schools and universities have autonomy. I think that it is good that they have that autonomy, want to make decisions and have the latitude to make decisions. I would not say that I think that there should be something that you would want to mandate, but perhaps there has to be some kind of formal conversation between, in particular, Ofqual and schools about the impact on assessment.

You asked about safeguarding. Again, I am thinking about this within the education and schools context. You have already pointed out that it makes a lot of mistakes. A lot of people have pointed out that it can go rogue and start to say some quite dangerous things.

I do not know whether there needs to be some kind of central set of advice. As I said, schools have autonomy. In lots of ways, that is a good thing, but there are important things about the impact on assessment and, as you point out, the impact on safeguarding and the dangerous things that it can say. We need to take those very seriously. They are not necessarily easy issues to address.

Professor Luckin: The one thing that we have to work out is how to change the balance of power. At the moment, the power is with the tech companies. We are here talking about this because a tech company has



HOUSE OF COMMONS

produced a technology and made it available to everybody. We have known that this technology has been around for several years. People like me—and Daisy, too—have been talking about the need to think about AI in education for a long time, but it took this moment to wake people up. We have to find a way of shifting the power. For me, the only way in which you can shift the power is through education—through really helping people to understand.

I have not had a chance to read the White Paper, but I know that there is an idea around a sandbox in it. That might be a really useful educational tool, if we can make it an educational tool. One way in which we might start to shift the power balance is by joining regulation and education together—making the way in which the regulation is developed an educational process, so that people start to realise why they need to care about this.

We will never, ever keep up. I agree completely with Daisy that prediction is difficult, but I can predict quite confidently that the technology will continue to advance. It will get better, but I agree that it will not be perfect and will do things that we are not expecting. We cannot predict exactly how humans will react to it. We can predict that a lot of people will want to make money out of it, that they will make a lot of money out of it and that many of them will be unscrupulous. We cannot predict precisely what impact that will have on people, other than that it will potentially be bad. We can also predict that many good things will come out of this.

We should look at the things we can be sure of and try to work out how we can educate people to be able to cope with the uncertainty of the things we cannot be sure of. I feel that having an educational aspect to regulation might be a way of starting to shift the power balance. As I said in response to Stephen's first question, we really need to educate everybody. Of course, that is a huge task, so we have to start somewhere.

Daisy Christodoulou: Can I come in on one other thing?

Chair: Very briefly. We need to get on to some other questions.

Daisy Christodoulou: I hesitate to say this because I do not think that it is that realistic, but the one thing that we as an organisation and a lot of organisations would love to know, particularly in the case of OpenAI, is how they have trained their model. They have not released that. They are called OpenAI, but what they are doing has a lot of closed aspects. That makes it incredibly hard to evaluate what it does and how it produces its results.

There is an element of AI that is inevitably a bit of a black box, where even the developers themselves may not know how it is producing its outputs, but there is an element of AI where they could be more open. They have published some aspects of how it is trained, but not the full



HOUSE OF COMMONS

details. I think that that will be a future area of regulation, not just in this country but globally. To what extent do you force organisations like OpenAI to open up and share how they train their models?

Q272 **Chair:** I note that one of the five principles that the Government have set out today for the governance of AI is transparency and explainability. If that is to be applied to regulators, we will want to go into some more detail on it.

Thank you for that. That was a useful snap reaction to what has been put out today. What we want to do for the rest of the session is to benefit from your understanding of how AI is currently being applied in schools and what seems to be around the corner.

Ms Christodoulou, can you describe the reality now of the deployment of AI in school settings, as you see it? Can we keep the answers brief? As you can see, lots of Members are eager to ask lots of questions.

Daisy Christodoulou: Sure. The reality of the deployment at the moment is described by the line, "The future is here. It is just unevenly distributed." There is an uneven distribution. There are some teachers and some students who have started using it a lot, and there are some who have not heard of it. It is spreading.

What are the areas where people are using it and looking at it? You have the problematic areas, where students are using it to cheat. We are already starting to see that that is an issue. You have issues to do with whether teachers can use it to plan resources and to create questions.

Q273 **Chair:** For those of us who are not within the education world, can you say how it is being used at the moment?

Daisy Christodoulou: By teachers?

Chair: By teachers and students.

Daisy Christodoulou: For teachers, you have someone coming up on the next panel. You have teachers who are using it to create resources. A lot of what teachers spend their time on is planning lessons and creating resources. I know that lots of them will use ChatGPT to help with that.

I will give one example of where I think that it works really well. It is really good at text summarising. If you take a complexish text where you would vouch for the accuracy—Wikipedia is a good example—pop it into ChatGPT and say, "Can you rewrite this so that it is appropriate for a 10-year-old?", it does quite a good job. That is quite useful. It can do things like that. With something like that, it tends not to make too many mistakes, either, because you are giving it the information. Something like that can be pretty useful.

I have also seen lots of teachers and educational organisations using it to create questions. This is where it gets more problematic. If you make it more open and give it more latitude, it makes a lot of errors. It makes an



HOUSE OF COMMONS

awful lot of errors. I have a little collection of them that we are storing up.

Chair: When you say questions—

Daisy Christodoulou: For students.

Chair: Questions as part of teaching or questions for exams?

Daisy Christodoulou: As part of teaching. For example, at the end of a lesson on Pythagoras's theorem, you want to give students a few questions. You may already have some that you have created, but it is always nice to have more and you may want to mix it up a bit, so you ask ChatGPT, "Can you give me a question about Pythagoras's theorem?" It will do it, but the problem is that it makes mistakes. I mention Pythagoras's theorem because that is one of the ones we have looked at. It keeps making a really baffling mistake where, essentially, it confuses 14 squared and 13 squared. Thirteen squared is 169 and 14 squared is 196. I am convinced that it is just mixing up the numbers.

Chair: It is dyslexic.

Daisy Christodoulou: I do not want to say that, but who knows? If you want to use it in schools, we are nowhere near just being able to automate and put it out there.

Q274 **Chair:** How widespread is this? You say that it can be used to generate questions at the end of a lesson, but how widespread would you say that its use by teachers is currently?

Daisy Christodoulou: I genuinely do not know. I would suggest that at the minute it is probably used by a small percentage of quite enthusiastic teachers, but I do not know. I have not seen any data on it. That is my feeling.

Q275 **Chair:** Professor Luckin, do you have a perception of what the current uses are and how widespread they might be?

Professor Luckin: Do you mean the current uses of ChatGPT?

Chair: No—of AI generally.

Professor Luckin: Okay. There is quite a long history here. There is a great deal of difference between what happens in research and what happens in practice. I am going to talk about what happens in practice, because I think that that is what you are interested in.

Chair: Yes.

Professor Luckin: Two characteristics of AI are super important here: adaptivity and autonomy. We can use AI to build adaptive systems that will help a learner to progress at their own speed, so to speak. There are a lot of them around. They might alter the difficulty of the task that a student is set in order better to meet that student's need. They might



alter the nature of the hints or tips that are given in order to support that student to be successful. Then they can give good-quality feedback to teachers.

Many of them have been around for quite a long time. There is lots of data to suggest that they are pretty good—not all of them, but the well-designed ones. Studies that have been conducted with those systems have demonstrated that they are as good at tutoring—not teaching—as a human tutor when working with a whole class, remembering that the AI is working one to one. They are not as good as a human working one to one, but how many students would get that?

Q276 **Chair:** You place an emphasis on teaching versus tutoring. Can you explain the difference?

Professor Luckin: Teaching is a far richer activity than tutoring. Tutoring is just part of it. However, that is very useful. If you have an AI system that can do part of the teaching job and that can give you, as a teacher, really good feedback about your students, then you know, as the human teacher, where to offer your support in particular. These are very useful technologies that have been around for a while.

Q277 **Chair:** I want to pause there for a second. Am I right in thinking that, in a mixed-ability class that was studying a particular subject, it might be possible to have different approaches tailored to either the particular level or the particular learning preferences of individual students for that common subject? Would that be accurate?

Professor Luckin: That is exactly what I am talking about. The important part is that you get feedback as a teacher and you get feedback as a learner. You can also get feedback as a parent. The British company CENTURY Tech is a very popularly used platform, not just in the UK. There are other companies, too. In the US, Carnegie Learning has produced these kinds of systems for several decades and has improved them all the time. We know that.

There are also ways in which AI can help teachers to track down the most appropriate resources for their students. That is a different kind of adapting—not adapting to the students, but adapting to the teacher’s needs for those students, if you see the difference. It is not directly teaching—it is just helping the teacher to find the resources that are out there and available and that best meet their needs. These kinds of systems have also been around.

We also have AI being used to analyse and interpret students’ facial expressions and voice data. That is mainly in the university sector, but we have systems that are being used and that are not just research tools. We know that data captured about facial expression, when combined with data captured about voice, is much more reliable than either of those modalities individually in helping you to understand whether a student is particularly stressed or anxious. You do not need a watch saying that



HOUSE OF COMMONS

your heart rate is going up. Therefore, if you are a counselling AI, for example, you can adapt what you are doing to that student's needs. There is a lot of stuff out there that is being used. Daisy is right to say that it is not evenly distributed, because it all costs money.

The other type of AI that is being used—although, exactly as Daisy said, it depends on the teacher and their enthusiasm—is just the normal tools, such as dictation. AI is really good at transcription these days—things like voice-activated personal assistants, where you get your students talking, maybe in a foreign language. Does the voice-activated assistant understand them? No, it does not. That is interesting. Those tools are readily available. Teachers who understand this a bit can see how they can integrate them into the lessons. Again, it is not evenly spread out, but it would be a mistake to think that it is not out there already. It is.

Q278 Chair: That is extremely helpful. This Committee has had a long interest in this. The predecessor Committee published a report on algorithms in decision making, so our inquiry is very much in a tradition of advice to the Government on this.

As a final word on this, given what you have described as the current deployment, is there any evidence that AI improves student outcomes?

Professor Luckin: That is a really interesting question. There is evidence that we can improve student outcomes with the kind of AI that is now being developed, which can track the learning process, not just the learning product. A lot of these are not necessarily quite available at scale yet, but some are just starting to come out. There is certainly a growing body of evidence that students who might have been falling through the net can be helped to be brought back into the pack, if you see what I mean. I am sorry. That is not very good language, but you get what I am saying.

Q279 Chair: No, it is very clear. I do.

Professor Luckin: It is not so much about whether the GCSE results are going up as about the fact that this student may now have a chance of sitting that GCSE, whereas without it they would not. If busy teachers have this technology available to them, they can make sure that students who have a particular need can learn at their own pace and that, as teachers, they get that feedback. That will get better.

Chair: Good. That is very clear. I will now go to my colleagues, starting with a former science teacher. Do you want to come in, Graham?

Q280 Graham Stringer: Very quickly.

That is very interesting. What is the quality of the evidence about improvement in student performance? Is that from controlled experiments, or is it just from observation and comparison with previous students?



HOUSE OF COMMONS

Professor Luckin: Most of the evidence that I would consider to be good quality is still in the research space. There may be some control trials among that. The reason I am hesitating is that they will not necessarily have been conducted in the classroom environment. Until that happens, you cannot be sure exactly what will and will not play out. It is one thing for something to work in a lab, or carefully curated environment, even if it is a school classroom, but often in these studies you do not see what is going on behind the scenes to make it work. We need evaluations of the tools at scale where we would be able to get the kind of quality data you are talking about.

We are starting to see data, but it is the independence of it that is really important. A lot of the data you see comes from people who are building the systems. I like to see some independent data as well.

We are starting to see that in research studies. That will grow, so expect to see more and more. As we get better at using AI to track the learning process, we will see even more important data coming out about how timely interventions in that learning process will produce much better learning outcomes. This is quite new and it is not out there at scale yet, but watch this space. I hope that helps.

Chair: Carol Monaghan, who is a former science teacher, has questions for both witnesses.

Q281 **Carol Monaghan:** It is utterly fascinating. I left teaching eight years ago when I became an MP. We were starting to see the basic use of some of these tools when I left teaching.

Exams have developed throughout history. One of the reasons exams were introduced was to level the playing field. Before that, people from privileged backgrounds were able to present themselves to a university and get into it. Exams allowed access to a much wider range of people.

Over the past 10 years we have become more aware that some young people will struggle and not reach their potential in an exam. We are seeing greater use of continual assessment to determine awards in national exams.

Having listened to what you have said this morning about the use of ChatGPT, for example, to come up with essays or assignments, do we need to have a good, hard look at exam or assessment methodology once again?

Daisy Christodoulou: I work in assessment. This is what we do; this is our day job. I think we need to take a good, hard look at how we assess. I think that ChatGPT has huge implications for continuous assessment course work. It is very hard to see how that continues.

I have heard a few suggestions about different things you could do, but some of the people making those decisions do not realise how powerful a tool like ChatGPT is and that it is capable of producing original and very



HOUSE OF COMMONS

hard to detect, relatively high-quality responses to any kind of question. They will not be perfect, but they will be good enough for a lot of students. Uncontrolled assessments, where you are not sure how the student has produced the work, become very problematic. There is value in having traditional exams. You know how the student has produced it; it is their own work and thinking.

It is very hard to spot ChatGPT tasks. We did some research. We got ChatGPT to write eight responses to a task we ran recently. This was an assessment of eight-year-olds' writing. I know they were only eight-year-olds; we will repeat it with some older students. We did an assessment with 50,000 eight-year-olds. Of those eight pieces, all the ones that we designed to be the good ones came out in the top few per cent. Essentially, they reached that ceiling.

We asked our teachers when assessing them to see whether they could spot them. They could not; they were more likely to pick an essay that had been written by a real child and say that had been written by ChatGPT than spot the ChatGPT essay. It writes very good essays that are very hard to detect by humans.

I have heard people ask whether we could get AI to check AI—or plagiarism detectors. The research I have seen is not encouraging. We tried some of the models that are out there at the minute. Again, they do not have high accuracy rates of being able to spot them.

These are the things we have to be looking at and thinking about. A few educational institutions are sticking their heads in the sand and wanting it to go away.

We have had a problem at universities for years with essay mills. Essay mills have been perhaps a relatively small percentage. I think people now pretend it is not happening. You will not be able to pretend this is not happening because it has the potential for 100% of students to be using it in very rapid time.

There is a question about whether it will be able to produce university-level essays. The point is that, even if it can produce them only at the 50 to 60 percentile, by definition that will be good enough for 50% to 60% of students and they could get some value from it.

The other point about it is that it is just so easy to type in a question and get a response.

Q282 Carol Monaghan: A few weeks ago I gave a keynote speech to an academic audience about the use of AI in healthcare. I decided to start this speech with a chunk from ChatGPT. I asked it to write my introduction and include a joke, which it did. I delivered the first two minutes from ChatGPT with a joke that they laughed along with. At the end, I told the audience that that was ChatGPT.

You could see the faces. There was not a sharp intake of breath, but



HOUSE OF COMMONS

clearly I had delivered something that was good enough for an academic audience. Quite disturbingly, I had a couple of my own jokes later, and they didn't laugh—maybe ChatGPT does better!

There is an issue in how we are assessing. We have talked about coursework and essays. That is a big issue. What about scientific subjects? What can it produce? If I set a physics assignment for my students, how will it cope with that?

Daisy Christodoulou: Some study data from GPT-4 came out in the past few weeks. People gave it a bunch of American grad school exams for a whole range of subjects: some science ones, medical ones, law and whatever. It aced them all; it did very well on all of them. For something like physics and things like that it is very good at solving those questions and answering those assessments.

Personally, I think our response has to be that we have to look at assessments in more controlled environments. I was an English teacher. I recognise that for some subjects that is not easy, particularly at university level where the model of going away and writing an essay is really important. I do not pretend this will be straightforward, but, as you pointed out, its ability to produce very good, fluent prose is very hard to distinguish. It is something you cannot ignore.

Q283 **Carol Monaghan:** You have also analysed ChatGPT's performance at assessing different pieces of work. How did you get on with that?

Daisy Christodoulou: This is where it is less good and this is the problem for us at the moment. I believe that all the things it is good at, at scale, in education would feel like less socially useful things. It is very good at teaching and giving students essays that are not their own; it can do it at scale and rapidly.

For the more socially useful things in education—I have talked about creating questions, but also marking—it is not as good as we had hoped. We have done the work to plug ChatGPT into our assessment systems. We can now quite easily and rapidly compare human and AI judging. It does not make the same decisions as humans; it makes quite odd and baffling ones. It is not making the right decisions at scale.

The other thing about it—this the difference between the old AI systems and the newer LLMs—is that the great advantage of previous and narrower AI systems, a lot of which we use in our back office, is that they are reliable and repeatable. You give them a question one day and they give an answer. You come back a week later and it will give the same answer. With GPT, that is not the case. In some ways, that is the strength of it, but for something like marking it is a weakness. You ask it to mark an essay one day and it will give it one mark; you come back five minutes later and ask it again and it gives something totally different.

Carol Monaghan: What about non-essay marking, again talking about science and maths?



Daisy Christodoulou: We have not experimented with it, but I have been talking to people who have. It does weird things. For example, you will get it up and running and think you have got it working; it will be marking some multiple-choice questions and giving them all a mark. Suddenly, it will switch and start giving half-marks. It does odd things; it is not reliable. Everyone keeps saying, "We haven't used it; we are waiting for ChatGPT-4, API key." For everything I say here, maybe tomorrow we will get our GPT-4 API key and maybe all of this will change, but where we are at the minute, it is not doing things reliably enough; it is not consistent enough and makes weird errors. There is no way you could use it even in a low-stakes system at scale for assigning marks to essays.

The other concern is that I am not even sure it can count. In the process of assigning a number to a mark, we are not sure how it is doing that. That goes back to some of the training issues. It might be a little bit better in providing a general passage of feedback. We are looking at that a little bit.

If I was to give it a grade, I would say it is absolutely brilliant at producing essays for students on their own. It is not very good at marking, and as for giving some general ideas for feedback I would say that maybe there is some potential there. I put that in as well.

Carol Monaghan: Of course, multiple-choice questions are the easiest for any teacher to mark; it's bang, bang, bang.

Daisy Christodoulou: We have built and have been using for a few years now a narrow AI system to recognise multiple-choice questions on a grid—ticks and crosses. That works beautifully. It is not very glamorous, but it works very well; it is very repeatable and cuts down teachers' workload. There are narrow uses of AI, which are not very glamorous but are very repeatable and transparent. They are fantastic. We have a mathematical model and can see exactly why it is doing what it is doing. Rose talked about some of those as well. As for those models, there is a lot of potential. At the minute, I would say that a lot of people working with GPT would say it has a lot of baffling things about it.

Professor Luckin: Something is worrying me about this. I think that is true for ChatGPT, but a different large language model trained on a more accurate dataset would do better at marking. I do not think that is the question. What you drew attention to was the fact that originally exams were introduced to try to bring about equality.

Carol Monaghan: We seem to have forgotten that in some of this.

Professor Luckin: Exactly. This is a fabulous opportunity, because if we now have AI that can pass our assessments, in the short term we have to do all this worrying about how to deal with the fact we cannot change assessment overnight, and how we are going to deal with this fact. But in the longer term, in truth, if I can have a system that does not understand



HOUSE OF COMMONS

anything and has no idea how any of this applies to the real world, is there not something wrong with the assessment in the first place?

Daisy Christodoulou: Rose, I would fundamentally disagree with that. I agree with everything you have said so far, but I think you are wrong on that, and I think it's dangerous.

Chair: Perhaps Professor Luckin can develop the point.

Professor Luckin: This is what it is all about. What does it mean to know something? Our education system may be called a knowledge-based curriculum, but it is not; it is an information-based curriculum. It is beautifully designed and I understand why people are so committed to it, but it will not produce people who really understand. This is what ChatGPT is drawing our attention to, because it collates much greater bodies of information than we will ever be able to collate.

Over the years, we have developed AI in our own image because we do not have another image of intelligence. We have only our own image of intelligence and so we build it in our own image, and we build it to do the things we value.

I think we now have to have a serious think long term. What is it to be intelligent as a human? What are the real characteristics of human intelligence that we want our populations to have? Surely, they are not the ones that we can easily automate; surely, they are the ones that we cannot easily automate. Believe me, there are a lot of them.

I am old enough to have worked in a computer science and AI department when IBM's machine beat Garry Kasparov at chess. I remember the excitement; people were cock-a-hoop, saying, "We've cracked intelligence; we can beat Garry Kasparov at chess—yay us!" with almost as much hubris as ChatGPT.

Over the coming months and years, people realised that some of the things we took for granted—vision, for example—was much harder to do. It has taken us decades and we are still not perfect. We thought chess was a signifier of intelligence. We need to look at what it is to be humanly intelligent. If we do not get that right, we will not work out the right relationship that humans must have with artificially intelligent systems, because we do not want to repeat; we want to use these systems to make us smarter. That is the most important thing.

Carol Monaghan: I am aware we could talk about this all day.

Professor Luckin: This is all about equality, which was where you started, because it might help us. If we get it right and use these great techniques for analysing learning processes and are able to highlight some of these uniquely human features of our intelligence, we might open up qualifications to populations that have previously been denied them. It is a fantastic opportunity, but we have to recognise we have to move on from where we are now.



HOUSE OF COMMONS

Chair: Ms Christodoulou signified dissent to something Professor Luckin said, so perhaps she could explain that.

Daisy Christodoulou: I agree with a lot of Rose said earlier and with quite a bit of what she said there; I just want to tee up the disagreement.

I absolutely agree that we need to be focusing on real human understanding. I also agree with Rose that I do not think that GPT has that. We need to focus on what it is that makes us human and get that human understanding.

I think the difference arises when you say that the things we need to be focusing on in schools are not things we can easily automate. I disagree with that for a couple of reasons. We want students to have the advanced skills that computers and AI do not have, but we also know that in order to get those advanced skills there is no shortcut. You cannot short-circuit it; you have to go through the basics. The student, unfortunately, will have to spend time grappling with the problems that AI can solve immediately and can be automated immediately, because they cannot get to the more advanced ones without that.

Chess is a fascinating case study, for a number of reasons. Everyone thought that Deep Blue and AI would kill chess; they thought that the minute a computer solved chess nobody would want to play any more. The reverse has happened: there has been a boom in chess; people love it. When you teach a child to play chess now, do you say to them, "Don't bother learning how the pieces move; you need to focus on the strategy and the things AI cannot do"? Of course not. For a student to become good at chess they have to learn the pieces first. Computers have affected chess education not by telling students not to worry about the basics and the things the computer can do; it is by developing some interesting new training patterns—for example, ChessKid.com—and ways of teaching chess that focus on the basics in a fun way. That is the impact technology has had on chess.

We cannot say to students, "Don't bother with the things computers can do." That will stop them getting the skills they need.

Just because a computer can do something does not mean there is not value in a human doing it. We did not stop teaching PE when the car was invented; we did not stop teaching drawing when cameras were invented. If ChatGPT can read better than a lot of humans, it would not stop us teaching reading, because you get a tremendous amount of value from reading. You do not have to be as good a reader as ChatGPT to get value from it.

The risk is that we just end up defining education completely instrumentally and saying that every time ChatGPT does something new we will change our curriculum and should not teach kids that.



HOUSE OF COMMONS

I absolutely agree with Rose that we have to focus on student human understanding and what that is, and perhaps the positives of GPT are that it will make us think more about what that is, but that does not mean abandoning the basics or fundamentals.

Chair: That is very helpful.

Professor Luckin: I never said we should abandon the basics or fundamentals.

Daisy Christodoulou: You said we should automate—

Professor Luckin: No.

Daisy Christodoulou: I wrote it down. You said we should not be looking at the things we can easily automate. I disagree with that.

Chair: We like spirited witnesses. Dawn had a question she communicated via our own system on the implications for more exams. Do you want to put that to the witnesses, Dawn?

Q284 **Dawn Butler:** How do we level the playing field? If ChatGPT can write essays really well, does that mean we will go back to having more intensified exams and an exam environment, which is what we were moving away from?

Professor Luckin: I really hope not.

Daisy Christodoulou: I think it does—and it should.

Q285 **Chair:** Let us hear from Professor Luckin first.

Professor Luckin: Exams are stressful; they do not produce accurate outcomes; they are outdated; they are not part of the learning process.

What I was trying to say—I did not say it clearly enough—is that what we teach children about the basics is fundamentally important, but it should be a tool to help them gain the more advanced human intelligence. Yes, you still have to do it.

Daisy Christodoulou: So we should be teaching things we can automate. Can we agree on that?

Professor Luckin: Yes.

Q286 **Chair:** One at a time; otherwise, it is difficult to follow it.

Professor Luckin: Through schools you reach higher education, and these should not be the things we assess. This should be the final part of education, because if we assess the more advanced skills we will in the process assess the fundamentals. We are just seeing it in a slightly different way.



HOUSE OF COMMONS

I spoke earlier about being able to track the learning process. You can track very nicely, using continual assessment, exactly how much someone is learning about the very basic things they need to understand. I could not agree more.

We do not need those final exams. Once we get our technology sorted—most importantly, once we get our data infrastructure sorted within our education system; the UK is good in this respect and we can build on what is there—there will be less need for these final exams.

I totally understand why Daisy says this is a way of stopping cheating. I get that, but it should be a short-term fix, not the long-term goal. The long-term goal has to be to sit down and look at what we really want people to be able to understand and do through our education system.

Q287 Chair: Ms Christodoulou, what is your response to that?

Daisy Christodoulou: You are absolutely right to make the distinction. We have an end goal and we can break it down into steps on the way and look at what those steps are. You can do that for lots of different subjects.

Often, the steps on the way do not look like the end goal. I think you need to assess both; you need to be assessing the end goal and those advanced skills we are talking about, like essay writing, complex maths problem-solving and complex science problem-solving.

You also need to be looking at the steps on the way that will get you there, because you cannot start assessing six-year-olds with those complex advanced problems; you have to break it down into the smaller steps. That is what a lot of the good AI systems do. They take the complex skills, break them down and think how best to teach and assess them.

That is what we do at No More Marking, the company I work for: big assessments of essays and complex assessments of advanced skills. We break them down into small steps and offer schools a multiple-choice tool to be able to design multiple-choice questions that will lead up to that end goal.

I definitely think you need to do both; you need to assess both and pick the right assessment and assessment structure for each. The coming of GPT means that doing any of the assessments in uncontrolled conditions just puts you in a situation where you do not know whether that work has been completed by the student. That is why I think you have to have exams.

Chair: Let us take some more questions from colleagues on this theme. It is important to say that our inquiry is not actually into ChatGPT as one particular tool; it is into AI generally.

Q288 Aaron Bell: At the risk of continuing this a bit further, is not the analogy



HOUSE OF COMMONS

with the time pocket calculators came in during the 1980s? That changed arithmetic fundamentally. They were seen originally as cheating and there was also an equity issue when tablets came in because they cost a lot of money in the first place, only some students had them and they were banned. Then we realised it was more sensible to do it because everyone would have access to computers and calculators when they were older and we did not do as much long division, or we did not use slide rules.

If the future is that people will use tools like ChatGPT, or whatever comes after it, to do creative work in their adult lives, do we not need a system that embraces that, but maybe not at primary school, because we do not let five and six-year-olds play with calculators to learn how to add up—is it all about the building blocks?

Daisy Christodoulou: What we need to think about is how we get students to have the advanced skills. I am completely with you and Rose that the aim has to be the advanced skills. The question is how we get them. What we know is that it is really important to have stored fluently in long-term memory a body of facts. Often, that is what allows you to do a lot of the creative thinking. On the “chess” point, you have to know how the pieces move. That is the point.

We still have exams for primary students that are calculator-free because there is value in getting to fluency. It is the same as reading a book. You can always look something up, not just on Google but in a dictionary. Dictionaries have existed for longer than pocket calculators, but we all know that if you understand 70% of the words in some text and you have to stop all the time to look up a word the fluency of your reading has gone.

It is not just pocket calculators. Any reference source is often fantastic for people who have that kind of fundamental knowledge and skills to be able to access it, but it does not help students who do not have those basics. I think that a large part of primary, even secondary, schooling has to be about making sure students have those fundamental basic skills and knowledge they need to be able to develop the more advanced skills, and to use all of those tools effectively.

Q289 **Aaron Bell:** In lots of ways it is about maths at secondary level in particular. I completely agree about building blocks, but maths is now a very different subject at GCSE from what it was at O-level. Is it not the case that that is how English, history and other subjects will change?

Daisy Christodoulou: I do not know to what extent maths is a different subject. I have a copy at home of the first exams in England in 1859 or 1869. You are right. The maths curriculum is quite different. I am not sure that all those reasons are to do with pocket calculators. There is a lot more statistics in modern maths, and I am quite keen on that, especially given the organisation I work for. We use a lot of stats. That has come about because we have seen it as being more useful.



HOUSE OF COMMONS

Does the curriculum need to change? We need to think concretely about how we would change it. I still think, as a former English teacher, that the skills of reading and writing will be phenomenally important at advanced levels in terms of the economic value they add, but also—we often forget this when we talk about it—in terms of just the personal value they bring. The ability to read a novel or poem and enjoy that is very important, too.

It feels to me that in English reading and writing will always be important; they are fundamental skills that underpin everything else. We know that in the UK, as in some other countries, unfortunately we have a number of students who do not have the literary skills they need to engage in the modern world. It seems to me that those are all really important things we need to focus on. We cannot get distracted by thinking we can forget about the basics because ChatGPT will do that and we can just focus on the higher-level stuff. You have to go through the basics to get to the higher level.

Professor Luckin: It is such an interesting question and it is a very complicated answer. Literacy and numeracy are fundamental; you have to have those in there.

The “calculator” issue is interesting because a calculator is not intelligent. We are talking about something much more sophisticated. If I look at a high-level skill—say, metacognition—we know that people with sophisticated metacognition do better in the world of work and in their learning outcomes.

You cannot teach metacognition in a vacuum; you have to teach it within a subject area, so that speaks to Daisy’s point about the need for subject areas. We also know that it is different at primary school because it is a developmental aspect, so the way you would go about teaching metacognition through a subject area—science, English or whatever—is different in a primary school from a secondary school. The point I am making is that focusing on metacognition rather than the subject and taking literacy and numeracy as given—we have to have high-quality versions of those—alters the way you teach and assess. That is what is important here. I am not saying we do not need those basic things; it is just that we need them in a different way, because we want to continue to be the most intelligent form of intelligence.

A very important issue that has to be grasped is the risk of offloading. Our brains are naturally lazy and, as is being shown with ChatGPT, if something else can do something for us we are highly likely to offload and say, “I don’t like doing this; it’s making me think hard. There’s this thing that can do this.” We do not want that to happen. In order for us to be productive it could be effective to offload certain things. The most important thing we need to decide is: what are the things we are happy to offload and what are the things we are not happy to offload? We need consciously to design systems that recognise that we will want to offload.



Aaron Bell: And fight against ourselves?

Professor Luckin: Exactly. We need to work it out. That is where AI, if we look beyond ChatGPT, can be super helpful. As I was saying to your colleague, we are now getting very good at taking data signals and learning about the process that a learner is going through.

When I was in further education, teaching adults with special needs, I would have loved to have known what was going on. I could have been so much better than I was.

You cannot cheat with that; that is your data. Formative assessment using AI is a way we can really move forward with assessment where it is less stressful, it is in the background, and you cannot cheat with it. There are ethical implications to that. I am not saying it is easy, but it is possible and it is not a long way off. We need to recognise that this stuff is coming and we can use it, and it will hit that point of equality if we design it in the right way.

Chair: Because of the human rather than machine interactions this morning we are running over time. I must ask my colleagues to be brief in their questions and the witnesses to be brief in their answers.

Q290 **Aaron Bell:** I will be very brisk, Chair. All witnesses in this inquiry have been reluctant to answer this question, but do you think that in 10 years' time education will look fundamentally different as a result of AI, or will it just be an incremental improvement?

Professor Luckin: Absolutely, it will; there is no question that it will look fundamentally different, but it is how it will look fundamentally different.

A few years ago I said I could see the dystopian and the optimistic version that I want. The dystopian version is that because it is cheap we could have a horrible future where most of the population are educated by AI teaching machines with some minders looking after them—it is cheap and we can track what people are learning, et cetera—and people who can afford it will have the rich human interactions integrated with AI and all the good stuff. That is horrible, but we are at risk of that happening.

What we have to do, but we really have to make the decisions now, is get over the fact that we have invested a huge amount of energy in something, our current curriculum, that is beautifully designed, but it has to change because the world is changing. If we want the optimistic version where we go back to the reason we had exams in the first place and want to bring about equality, we have to look at the situation differently.

Daisy Christodoulou: I agree about the dystopian vision. I think the thing that a lot of people worry about with AI education is: will it just be kids sitting in front of screens with eye trackers and rewards if they get certain questions right? I agree we do not want that, but I am fairly



HOUSE OF COMMONS

optimistic about it, because not only does that sound not very nice; I do not think people and kids want it. I think we saw in the pandemic that if that was going to happen it would have happened, but it did not because people wanted to get back to school.

When you ask what education will look like in 10 years, I think it depends on how you define “look”. I think it could look quite similar in that I think the physical aspects of schools are important and the physical place matters. There is a way in which schools might look very similar to schools today, but what goes on around them and how everything works around them could have been transformed.

The analogy I often like to use is sitting down to eat dinner. There are aspects of that that are very similar to sitting down to eat dinner maybe 100 years ago, but the food on your plate and everything around it has been completely transformed by technology. That would be my analogy. Things might look the same, but technology, in all kinds of ways, has the potential to transform it for the better.

Q291 **Dawn Butler:** I have a quick question on the difficult subject of bias. Where do the AI tools developed by edtech firms source their data?

Professor Luckin: It depends on the AI company. That is not meant to be a facetious answer—it is a really good question. It is a question that anybody who is thinking about buying any of those systems should be asking. Where is your data coming from? What are you doing with that data? Explain it to me in a way that I understand, and do not tell me it is commercially sensitive and you cannot tell me, because that is rubbish; you can. That is a really important question.

We have to know where the data is coming from, how reliable it is, how representative it is for the population that will be using the system and how much it is updated. On the big point Daisy made earlier about not knowing what OpenAI is doing with the data it has, we know nothing about what it is doing with the data it is collecting every time anybody types anything into it.

The same question arises: what happens to the data that is being put into these systems? Does it add to the pool of data from which the system learns, in which case are we increasing the bias because of the people who are using it? Do you see what I mean? It is a really important question. We have to empower people to ask that question and know how to ask it.

The Institute for Ethical AI in Education, which produced a report a couple of years ago, tried to help educators to understand the questions to ask when procuring AI, and some of that was around data. You are right; it is really important.

Daisy Christodoulou: A lot of the narrow AI—the applications—will be coming up with its own sets of training data, whatever it is for. Often, the



HOUSE OF COMMONS

applications will be using human judgments that they have gathered previously to help predict or make judgments for the model.

The issue with GPT, as I said earlier, is that we do not know what it has been trained on. That is a huge issue for working out anything to do with bias. A lot of the narrow AI models might not be completely transparent; there is an element of opacity in all of this, but you should be able to see what they have been trained on and, therefore, how they are making their decisions.

I definitely come back to what I said before. One of the key things about regulation will probably be the transparency of the training data.

Q292 Tracey Crouch: I apologise for being late. It has been a fascinating session.

Professor Luckin, you said that we need to be, or we want to be, the most intelligent form of intelligence, which is a really interesting phrase. Do you think we can be?

Professor Luckin: I do, but we have to make the right decisions. The brain is amazing; there is so much of it we are learning about now. It is so far beyond what AI can do. Every time we think we have got closer to artificial general intelligence there will be some aspect of what we can do where people will say, "Oh, yes; there's that, too." I do think we can.

It is interesting that a lot of the regulators and large organisations, which are making an awful lot of money out of AI, say, "Don't worry. We'll have a human in the loop; it will all be all right." What if that human does not understand AI? If we do not help the human to get to the place where they can be a useful human in the loop, they are not that useful. I hope that is helpful.

Tracey Crouch: It is very helpful. I request not to be that human.

Q293 Carol Monaghan: We are talking a lot about where we are here in the UK. Have interesting aspects of AI been used in education in other countries? Where should we be looking to get a good or bad example?

Professor Luckin: Look at China, perhaps; that is fairly obvious but scary, but, in all fairness, there is a lot to be learned from that. Obviously, it is the country in the world that is doing the greatest amount of surveillance and it is very advanced. They know such a lot about their population and over time they will be able to track exactly who knows what, where and how well, and precisely what they are doing with it, and so on. Obviously, it is not the way we want to go.

I think we need to look at that and learn how not to go about it, but also to learn the power that data can bring with it. We need to look at a much more conservative way forward. The US is very good at investing in AI and education in a way that we are not. We are good at AI in the UK; we



HOUSE OF COMMONS

are certainly in the top five for sophisticated understanding of science and the development of AI systems.

The US is good in the amount it invests in it, but there are other much smaller ones—Scandinavia and parts of eastern Europe—taking a step back and thinking about how to help everybody get an understanding of AI. I loved the Finnish 1% project. Let us help 1% of the population to understand AI and then they can help another 1% to understand. If that had been here we probably would have said, “Let’s have 69% of the population understanding it.” That is completely impossible, but 1% is so achievable.

Daisy Christodoulou: I have just returned from Australia.

Professor Luckin: That is a very good example, yes.

Daisy Christodoulou: They have just set up a federal organisation, the Australian Council for Educational Research. It is doing some interesting things. It is a federal group and it can look across the states; it has a bigger remit. It also has links with other government departments. It has other conversations; it is not just an issue of education. It is thinking about it in other areas. It is doing some interesting things and thinking about this in some interesting ways, too.

Q294 **Chair:** I am intrigued. Ms Christodoulou, your company is called No More Marking. Why is it called that?

Daisy Christodoulou: Because in assessment terms the technical meaning of marking is applying a number to a piece of work. What we do—it does not involve AI—is use an approach called comparative judgment. It is human judgment, but it is human comparative judgment. The human will read the writing, but instead of applying a mark to one piece of writing they will read two pieces as a comparative judgment and will say which is the better piece of writing. They make a series of decisions like that. At no stage are they ever applying a mark to a piece of writing, but when all the decisions are crunched together all the pieces of writing will be allocated marks.

Chair: You have satisfied my curiosity admirably. I thank our witnesses, Professor Luckin and Ms Christodoulou, for a fascinating session. You have exposed a lot of the debates that we have in here, so we are very grateful.

Examination of witnesses

Witnesses: Dr Matthew Glanville and Joel Kenyon.

Q295 **Chair:** I introduce our next panel of witnesses. Dr Matthew Glanville is head of assessment principles and practice of the International Baccalaureate, the IB, which provides education programmes to nearly 2 million students worldwide. Dr Glanville is a former maths teacher. He joined IB in 2014. Prior to that, he held roles at Ofqual, the qualifications,



HOUSE OF COMMONS

exams and assessment regulator, and he has a PhD in mathematics. Welcome, Dr Glanville.

Joel Kenyon is a science teacher and is community cohesion lead at Dormers Wells High School in Southall, London. His website “inquestion” covers content related to but outside the national curriculum to help enrich lessons.

We are very grateful to both of you for coming today. Perhaps I may start with a brief question to Dr Glanville.

Given the publication of the White Paper today that places emphasis on existing regulators rather than on a new one, as someone who was formerly in Ofqual does that seem sensible to you, or does your experience lead you to favour another direction—in other words, a specialist regulator?

Dr Glanville: It is really important that people who have a fundamental understanding of education look at what artificial intelligence means in that context, rather than the other way around. I definitely favour existing expert regulators feeding into how it works in education.

Q296 **Chair:** That is very clear and helpful.

Mr Kenyon, thank you for taking the time away from your students today; I hope they will forgive our claim on you. We are all very interested to know how you as a teacher use AI in your classroom with your students.

Joel Kenyon: In particular, it is a resource. That is mainly how I use it, because in education and teaching a lot of time is spent on making resources. I just try to find ways in which I can shorten the time it takes me to make something. Sometimes I know what I want to make and how I want to make it, but it is getting that first step—for example, with a WAGOLL, what a good one looks like, or a WABOLL, which is what a bad one looks like—which would be an answer to an essay question. From that essay question there would be a generated answer and then they would critique it and say what is good or bad about it. I can also make it put in spelling mistakes. I will put in five spelling mistakes and the pupils have to go through and highlight those mistakes.

Q297 **Chair:** You used the expression “WABOLL”. I am not as familiar with that as perhaps other members. Just say what it means.

Joel Kenyon: What a bad one looks like; it is an essay that is not very good. Then, pupils have to critique what is good about it, and ChatGPT is really good at making a bad version of something. It can also make a really good version of something that the kids can then critique.

Chair: That is a WAGOLL, is it?

Joel Kenyon: Yes.

Q298 **Chair:** That is an important piece of learning for us.



You apply that. Is this a kind of personal thing? Does your school do it in a structured way?

Joel Kenyon: We centralise all lessons in science. All the different topics were divvied out, and we are all responsible for the topics that we made. All the resources were made as a department and we shared them all. When I was making mine and ChatGPT came on to the scene, I just went, "Oh, maybe I can use this," and then I used it to make mine. Some of my resources for key stage 3 and the personal ones that I have for key stage 4 and key stage 5 were made by me. It is not necessarily something the school is using, but it is something that individual teachers are using in order to speed up a process.

Q299 **Chair:** I see. You have the "inquestion" website. Is that promoting what you are doing in your practice more widely or making it available more widely?

Joel Kenyon: It is just making it available and highlighting it. It is a practice called hinterland, which is the stuff that is not on the curriculum that can sometimes make the curriculum a little bit more interesting—stories from a wide range of fields that are just a little addition that can sometimes make a subject that is quite dull or not that interesting to a pupil a little bit more interesting. I have written some articles on there that would potentially be interesting for another teacher to use and mention in the classroom and apply context or real-world applications to some of the science that we talk about. I also go through how you can use it in the lesson.

Q300 **Chair:** I see. That is very helpful.

Dr Glanville, the IB hit the headlines a few weeks ago by making a statement that the organisation will not ban the use of tools like ChatGPT for the IB exam. Will you explain what that means in practice and why you have come to that decision?

Dr Glanville: The IB's approach is that students who use passages of text or other things generated by artificial intelligence must reference that in the same way as they would reference any other material they have taken from a different source. That means that they need to put it in quotation marks, say after the part, "This is what I have used and this is where it has come from," and then include in the bibliography what system they used and what prompt they used to generate it.

The principle is that we are trying to make sure that students are transparent about the work they have produced and what they have referred back from somewhere else, whether it is an artificial intelligence system or another source altogether. This was published on 27 February, and on 16 March we followed that up with a new academic integrity policy that sets out a whole annexe dedicated to the use of artificial intelligence in this context.

Chair: I see. This is for assignments that are outside the exam room.



Dr Glanville: Yes, this is really for the coursework element that we have spoken about earlier. Obviously, with exams the students would not have access to the technology to do this, but we have coursework that counts towards the final grade.

Q301 **Chair:** I understand. Your requirement is that if AI is used you have to reference it and cite it. You were sitting through the previous session. You would have heard that it is very difficult to detect whether it has indeed been used. What is your thought process around detecting things where people have not made the citation that you say you require?

Dr Glanville: This is where the IB approach to that coursework is really helpful because we have always required that teachers are alongside the student during the process of writing their coursework. They need to have meetings with them as they are starting to think about what the coursework is and then working with them during it, looking at early drafts to the final place.

In that context, teachers are able to recognise when suddenly a whole pile of work that the student had not previously even thought about appears and can have those conversations with the student about where that came from and what that means. That is actually exactly the same approach we use with essays bought off the internet that teachers can recognise when a student suddenly goes from having no understanding to some understanding.

Q302 **Chair:** I see. Do you have rules that they sign up to in order to spot that and intervene?

Dr Glanville: Yes, we have clear instructions on how teachers should support students during their coursework, and teachers are expected to sign a declaration that to the best of their knowledge this is the student's own work. That has been in place for multiple years in the IB.

Q303 **Chair:** What has been your experience of policing that? Obviously, there is pressure on schools, as there is on students, to get good exam results. Have you found that you often have to audit the performance and practice of teachers?

Dr Glanville: We have found that teachers are very good at spotting where there is a step change in their students' learning that seems odd and are very much able to say, "There is something odd going on here." We have on occasions needed to be very supportive of our teachers saying that is a clear responsibility, because, obviously, parents and students can put a lot of pressure on teachers to say, "Well, go on, look the other way." We are lucky that we have a very committed group of teachers who are prepared to stand up and say, "Actually, no, it is not," and the IB can then step in to support the teachers and the schools when they say these things.

Q304 **Chair:** Do you sample particular scripts or contributions and interrogate them, and go back and look at the history of the teacher's intervention?



HOUSE OF COMMONS

Dr Glanville: Yes, we do both. When we have our school inspection visits, that is part of what we look at. We also take a sample of all coursework, and review it and interrogate it. That is a very powerful tool of plagiarism, but, as the previous witnesses said, we do not think that is going to be the right tool to use for tests checking artificial intelligence because it is not going to be able to detect it in the same way as a high-quality essay mill essay cannot be detected because it is a unique piece of work. We do not know the students anything like the teachers know students, so that is where the best opportunity for identifying this comes from.

Q305 **Chair:** I see. Mr Kenyon, do you teach the IB in your school?

Joel Kenyon: No, I do not.

Q306 **Chair:** GCSEs and A-levels?

Joel Kenyon: Yes.

Q307 **Chair:** Is there an opportunity for people to use AI tools like ChatGPT in coursework for the exams that you teach?

Joel Kenyon: I teach science. We are 100% exams. There is the CPAC element in year 12—the practicals where they all have to produce a write-up. There is an opportunity for them to use AI to answer their questions or to write an entire report about an experiment that they have done.

I do not actually think it poses that much of an issue because all those CPACs have been done hundreds of times by hundreds of different students, and there is a wealth of resources already online that the kids can already copy, paste and edit slightly. They can use AI to help them answer their question, but it will probably just be easier for them to use a resource that has already been made.

What I do as a teacher and what a lot of us do when we look at the CPAC write-ups is say, "Did they do this in the practical session? Did they get the data in the practical session? Are the conclusions ones that we have spoken about or ones that would pop up? Do they fit where they are in the curriculum? Are they mentioned in stuff that is in topics 6 or 7 or in topic 4?" We take all that information in when we look at the work that they have given us.

Chair: It is a similar process. There is strong personal teacher scrutiny of what is being submitted.

Joel Kenyon: When we talk about the dystopian opportunities where the kids look at a screen, the problem with AI in order for us to detect whether a pupil has cheated or used AI when they are given coursework—the only way we are really good at doing that—is the relationships that we have with the pupils over time and the understanding of where they are in the curriculum. If you take out the human element, it makes it impossible to detect whether a pupil has



cheated. Therefore, a human element is needed because it is the only way we can figure out whether a pupil has used AI or not.

Chair: That is very interesting.

Q308 **Aaron Bell:** Dr Glanville, I appreciate the difficulty you are in with ChatGPT and all of this. If your method so far is basically that the pupil has advanced far beyond what we would expect or they have done it too quickly, surely ChatGPT can produce an essay in the style of the previous essays of that student. That is just a matter of laziness rather than trying to get a better grade. Is there not a real fundamental concern for continuous assessment and coursework, as we heard in the first session, that it is just not going to be possible once people start using these tools in a “smart” way?

Dr Glanville: I absolutely agree with you that there is a real challenge. As my colleague here has indicated, I think that teachers are able to understand what their students are capable of doing. It is a very interesting question. If they are using artificial intelligence to create what they could have created themselves, does that then become an assessment issue or a behavioural issue because they will get the same outcome? More generally, yes, I think it raises some really challenging questions about how we do assessment outside a tightly controlled environment.

I personally believe, and the IB believes, that there are some skills that really are not best to test in an exam context, and that is the problem. What is it that we value? You got a flavour of that earlier. What is it we think is important and what is the best way to assess that? It may well be that we cannot do some of those meaningfully in an exam context, and therefore it is better to assess them in the best situations we can get rather than not assess them in a more secure system. That is the tension that we deal with every single day.

Q309 **Stephen Metcalfe:** Based on your experience, how many students do you think are, first, aware that AI is available to help them; secondly, are using it; and, thirdly, how do we keep track of that?

Dr Glanville: The third part is the bit that would require actively going out there and doing high-quality research, which is not being done at the moment. My impression is that we have been talking a lot with our teachers through multiple routes—seminars, planned conversations to explain, webinars about how assessment works and conferences—and probably because of the statement most of our teachers are now aware and are starting to think about it. They are giving us feedback that their students are also aware of artificial intelligence.

In the short term, there will be a lot of students who understand and know about it. The challenge we have is how we turn that knowledge into ethical and safe use rather than using it in shortcuts, which has been described earlier.



Stephen Metcalfe: But that is just within the IB sphere.

Dr Glanville: That is just within the IB sphere. I have not had contact with teachers outside the IB.

Stephen Metcalfe: Would you care to speculate, or shall we ask Joel if he has any insight?

Dr Glanville: I think I am going to have to pass on the impossible question, and you are going to say you do not know, either.

Joel Kenyon: I do not think you will ever know. I do not think you will ever know the number of pupils who are using it. They are using it. There is no two ways about it. Some of them will say that they use it or will say that they have looked at it or made it try to do something, but I do not think they will ever admit to saying, "I used this to write this essay," because that is just not going to happen.

Therefore, trying to collect data to find out how many people are using it is nearly impossible because it is just a questionnaire, and you can just tick, "No, I do not know how to use it."

Q310 **Stephen Metcalfe:** From your personal experience, the students you are interacting with are aware of its existence and they are using it—possibly.

Joel Kenyon: Yes, possibly.

Stephen Metcalfe: But you do not really know in what ways.

Joel Kenyon: No, in terms of homework and stuff, we use an online system called Educake that does our homework for us. I know a lot of other departments do that as well. Where we are asking them to do essays—in English, for example—I can imagine it is going to be quite difficult for them to get feedback for their homework because we do not know if they have used a system or not.

Q311 **Stephen Metcalfe:** This is not in the area that I was going to question about. The explosion of AI is now turning into an arms race or a battle between you as educators and teachers, and big tech giants helping students combat you. It sounds like some bizarre video game. Do you think there is any responsibility on the tech companies to embed data within so that you cannot just copy and paste an essay you have written and you would have to retype it, in which case you might have learned something?

Dr Glanville: I would like to rephrase your initial premise. It is not about a battle. Artificial intelligence is going to become part of our everyday lives. That question of whether we get a dystopia or a beautiful future is important. We really need to make sure that we support our students in understanding what ethical and useful approaches are.

I would be grateful to the tech companies if they could do more to try to embed things, but I am also realistic that that is not a silver bullet that will address all of it. There will be other systems that do not have that



set-up, so even if the big ones act responsibly there will be others that still pose the same challenge. Let us address the real issue and not try to fight against something that will become part of everyday life in the not-too-distant future.

Joel Kenyon: I am going to echo that. It is going to happen. You could regulate it and say that you cannot put in large volumes of text and ask them to do it, but there will be another one using a VPN to access in a foreign country that does not have that regulation in place, and they will just be able to use it. You could ask certain companies to limit the amount of information it can give in or how much it can output, but, again, there will be another company that will just not do that. You have the big ones like OpenAI. Google has just released Bard. They are the big ones, but it is the smaller ones that you will find near-on impossible to regulate, so instead of trying to fix it with regulation try to look at the outcome and how we work with those outcomes rather than trying to change everything.

Q312 **Stephen Metcalfe:** That is very helpful. Thank you. I suppose that highlights that it is a powerful tool and it is here to stay. We interact with AI every single day of our lives now through our phones and decisions made on our behalf. Some of those decisions are very serious and some of them are not. Therefore, we judge the risk associated with individual artificial intelligences. Where would you place AI in education on that sliding scale of risk? Is it a big issue that we have to get our heads around, or is it just an interesting sideshow that teachers and educators will work out how to deal with?

Dr Glanville: I think—because I believe that it is going to become part of our everyday lives—it is really important, otherwise we are going to end up with an education system that prepares students for yesterday's issues, not tomorrow's issues. That is not only true of artificial intelligence; that is true of all of us. It is something that we really should be thinking about right now, so it is quite high up that list.

By working with our teachers, it does not need to be the showstopper to assessment. It presents some real challenges and makes us rethink how we assess and what we assess. That is the really important thing. Understanding the impact it is going to make on the world—how we need to teach our students to work within this new paradigm—is what is really important here, and that is a big challenge that we need to address sooner rather than later.

Joel Kenyon: Where we are talking about AI changing the face of education, I do not necessarily believe it is going to be a massive silver bullet that will have massive ramifications for how we teach. I agree with Daisy; education will probably just look the same. I do not think there will be large changes.

One of the things that we need to look at is the same issue that was raised before—where it is getting its information from. If you are a pupil



who is being asked to critically evaluate something, you want your starting base, which is what a lot of people use ChatGPT for. If it puts you at a starting point with a heavy bias, that changes the entire way that you write the rest of your essay. We need to know where they are getting their information from. ChatGPT, which gave you your false profile, is using data that is two years old. Google is using data that is immediately on the internet, and we already know the level of misinformation on the internet. The problem is that, if people are using AI in the classroom or using AI to answer questions for homework, it may inadvertently introduce them to misconceptions and biases.

Q313 Stephen Metcalfe: In terms of the risk to students' education, is the emergence of this technology low risk, medium risk or high risk? Can it be managed?

Joel Kenyon: For my subject in particular, I think it is relatively low risk, mainly because of the things that I teach and the way it is assessed. Because it is assessed in an exam, they do not have access to any of this. For the assessments and the questioning I do in lessons, they do not have access to them either.

I imagine subjects that are coursework-based will be significantly more impacted. For my subject, I think it is relatively low, but in ones with coursework I can imagine it is relatively high.

Dr Glanville: I agree with a great deal of what was said there. It is going to have a far bigger impact in some of the coursework areas. It is not a new challenge. We have always had the essay mills and everything else, but it brings it to a far larger part of the community. I would probably put it in the medium risk category.

Stephen Metcalfe: Excellent, thank you both for answers to the exam question.

Q314 Rebecca Long Bailey: We have spent quite a lot of time discussing the challenges of AI in education. What guidance or support would you like to see from the Government and regulators on the use of AI in education?

Dr Glanville: The IB is an international organisation, and therefore we believe that we have that obligation to support our teachers and students in understanding that. While I would be delighted to understand what the feedback from the Government is and the ideas from Government and the regulator, Ofqual, we also need to be doing that internationally so that we get the best possible thing. It is really important that we get information from somewhere to our teachers and our students soon.

The feedback I am getting from a lot of our teachers is, "What is working well? We have heard about this. How should we use it in our classrooms? How can you share good examples with us?"

The IB is trying to do that. We have a system called IB Exchange where teachers can exchange ideas, and artificial intelligence has been a big



HOUSE OF COMMONS

part of that in recent years. We really need to be taking the good examples we have and sharing them, and sooner rather than later.

Q315 **Rebecca Long Bailey:** Are there any particular countries—you work internationally—where you think they have the support and regulation right on this so far?

Dr Glanville: I do not think there is any particular country around the world that we are using as our benchmark. A lot of countries are thinking about this. A lot of countries are having some very sensible ideas about it. It is about taking all of those together and bringing them into one place.

Q316 **Rebecca Long Bailey:** Thank you. Joel, what regulation and support would you like to see from Government?

Joel Kenyon: It is mainly advice and signposting to what you can use it for. I do not think it can be regulated away. If you regulate it so much it will be unregulatable, it just will not work in practice.

The Government carry out teacher workload surveys and provide teacher workload guidance. The AI should be part of the teacher workload guidance in a way that people can use it to reduce the level of workload.

In terms of students, we should highlight the issues to them. The JCQ regulates and conducts exams. It uses all the information it gets from the different exam boards. Maybe we should use that information to say, "When you are looking at coursework, these are the things that you should look out for." Could the Government work with organisations to make a better plagiarism checker? That may or may not be impossible.

The guidance needs to be: how you can use it to reduce your workload; the ways that we should understand how pupils are using it to signpost the issues and how to deal with those issues; and best practice when speaking with other educators to try to find out how to limit the impact it has on schools.

Q317 **Rebecca Long Bailey:** Are there any specific skills that you think students might need in the future in education as a result of AI's growth? What sorts of things should educators be trying to teach their students at this time?

Joel Kenyon: I have a particular issue when we talk about the new skills for the future because all the skills that develop in the future were not taught in schools.

It is a baseline foundational knowledge that needs to be focused on. Do I think AI forms a baseline fundamental knowledge? No. It is a tool; it is not a piece of knowledge. I do not necessarily believe that the kids need to know how it works or what it does because they can figure that out quite quickly. They do not know how a calculator works, but they can use a calculator. I do not think this needs to be prescribed on the curriculum.



I do not necessarily believe it should be anything taught. It just needs to be mentioned in lessons as a tool that you can use if you wish to use it. Pupils are going to use it, but maybe we need to think about the best way pupils can use it, and that is how it should be fed in. It should just be, as when we mentioned calculators, "This is a tool that you can use. It is incredibly powerful. It is a tool that you can use to make your life easier or to make writing something a little bit easier." Do I think it should be taught? No, not really.

Dr Glanville: The really important point here is understanding of bias. The difference between getting something up on Google and getting it up with an artificial intelligence engine is that it sounds very authoritative, and it is not. All the previous conversations about bias and the inherent issues around what source material it learned from are absolutely critical, and that feeds into the principles of being able to understand and listen to other people's arguments and appreciate that there are different points of view, all of which is absolutely critical to education. That is part of what my organisation strongly believes.

That is the key for me. They need to learn not to take it at face value, but to interrogate it, to think about it and to look at it. Those are skills that they should apply to any piece of information they get from anywhere else. It is not a new skill, but it is just so much more important now because there is no filter to it. That is what we need to be aware of.

Joel Kenyon: They need to be aware of this. It is a new, additional one that they need to be aware of. We have been using the internet for years. Pupils have been using it. They need to learn about the bias from the websites that they get their information from. Bias is just another one of those things that we as teachers need to teach pupils to understand. This is just another one of those things that we add into that mix.

Q318 **Tracey Crouch:** I want to follow on from the bias aspect. It is a two-way process, is it not, from teachers and students? We were having a conversation about the bitterness that some of us still feel, decades on, from our friends copying essays days before the internet exists and getting better grades than us because they always got better grades than us. How do you as teachers respond to the bias that might have been generated in what is being given to you? You will have your own bias towards what a student is presenting. That student is always given something that is good, so how do you remove that to sense-check whether it has been AI generated or not?

Dr Glanville: I think what you are talking about is how we make sure that teachers are not biased and do not stereotype students. If that is the case, you are absolutely right. That is one of the key skills of being a good teacher—to be able to recognise your own biases and the baggage you bring to a situation and then work to remove them.

That is part of teacher training. That is a long way from conversations about artificial intelligence. It is absolutely critical that if you have



HOUSE OF COMMONS

teachers who bring their biases and impose their views on their students we deal with that with teacher training, no matter whether it is IB teacher training or the UK or anywhere else. Joel will be better placed to respond, I suspect.

Joel Kenyon: I do not necessarily think I am better placed. We all have a bias. The bias is there. Teachers have a bias. When we do our training, we are taught how to try to eliminate that bias or try to be the devil's advocate in every situation.

When pupils talk about creationism, or when I teach evolution, do I have a bias? Of course I have a bias. People who are religious would have a bias. We then form an argument between the two. We look at the pros and the cons from both situations, and then we discuss that as a classroom so they understand the biases. It is just a skill that teachers have.

Regarding copying the essay from the other person, if I was to have a student who just appeared in my lesson one day and gave me an essay that was made with ChatGPT, it would probably get a good mark, and I probably would not suspect it. As teachers, we—for some cases—work with them for years. For year 12 students and year 13 students, sometimes we have seen them over seven years, and we meet them multiple times and we have other teachers in the same department who understand those pupils as well.

We have conversations in the classroom where we say, "You had them in year 11. Is this the type of work that you would expect from them?" It is a yes or a no. When it is a no—it has seemingly come out of nowhere and there has been a massive jump in just a year—that is when we look at it and say, "Where did you get this information from? Is this plagiarised from someone? Did you get it from somewhere?" Those are just the conversations that we need to have. When they copy off others and change the wording and stuff like that, we are really good at noticing. We can always tell when they follow the same theme or there is the same understanding. Teachers are taught and teachers understand that kids are going to do it, so they need to learn how to spot it.

Q319 **Tracey Crouch:** What I am trying to suggest is this. Let us say that Dawn is an A-grade student, which does not surprise me, and I am a C-grade student, which would not surprise any of my previous teachers. We have both used AI to come up with very similar essays. Dawn's is considered A-grade because Dawn's has always been A-grade. Despite the fact that we have used exactly the same level of AI to generate the same answers in the same format, you are saying that I would also get an A because Dawn has got an A, or would you suggest that even though we have used the same tools I would still get my C? That is the challenge for teachers.

Joel Kenyon: You can test a pupil's understanding by what question they ask. Did you ask ChatGPT the right question? Did you ask the AI the right



HOUSE OF COMMONS

question? It would be the difference in the level of knowledge between you two where the outcome would be different. The output from the AI would be different based on the information that you gave it to begin with. Maybe that A-grade student would be able to notice where it is a little bit too advanced or it is mentioning something that they do not know and therefore able to edit it, whereas the C-grade student may not be able to say, "That is too advanced for me," because they do not know where the level of advanced stops because they are not there yet.

There is a bias where you have been getting As all the time, so I want to give you another A. You have been getting Cs all the time; maybe you have just made a good essay. It is a conversation. I am not just going to give it back to you with no feedback and blank it off. I will say, "This is a good essay, but I am also concerned as to where you got your information from." In a cold-cut scenario where we mark it and give it to them and that is it, yes, that can happen, but we mark with conversations and feedback to help build the students' understanding as to why they have got that mark, and then we also find out whether they have cheated.

Q320 Tracey Crouch: Do you think this is putting additional pressure on the teaching profession in terms of fact checking and everything else?

Joel Kenyon: It will. That is one thing that the Government need some guidance on. How do we do it in the best way possible that does not increase the amount of workload? It does not that much now, but I think it will in the future.

Q321 Tracey Crouch: I want to give you an example of the new Google AI function. I do not know why, but a friend put into it, "Is Tracey Crouch a good person?" You will be pleased to know that they cannot make moral judgments for themselves. Rather interestingly, it says, "She has also been criticised for her support of fixed-odds betting terminals," which is the complete opposite given I have resigned from Government over fixed-odds betting terminals. That means that AI—you talked about misinformation on the internet—has gone on to the wrong piece of information. If you as a teacher for whatever reason were asking your class to do an essay on me, you would now have to go back and check that.

Joel Kenyon: This is where subject knowledge is incredibly important. I know my curriculum. There will be parts of my curriculum that I do not know as well as others. I know my curriculum, and teachers know their curriculum. When you start, you do not know that much of it. You only know what you got from your degree. Once you have been teaching for a few years, you start to learn more and you find out what the misconceptions are. That is a test of subject knowledge.

If I was a politics teacher who was talking about you and what you have been doing, I would know and I would read and I would understand before I taught it, so I would instantly be able to pick up the



misconception. If I am teaching science and I ask them the structure of the earth and they mix up the order, I can instantly figure out that they have mixed up the order because I have good subject knowledge. That is one thing that is important in order to help combat the AI: teachers need good subject knowledge. That is part of teachers' standards, but it needs to be emphasised a little bit more.

Q322 Chair: That has a very important implication: you need to have teachers with specialist subject knowledge. As part of our diversity in STEM inquiry, one of the things we established is that in STEM subjects there is a shortage of qualified teachers, so that adds to what has already been a problem. You are a STEM teacher. Would you agree with that?

Joel Kenyon: There are two problems with this aspect. You have teachers who are leaving the profession in a short period of time, which means they have not gained the subject knowledge, and then they are replaced by a teacher at the very beginning of their career who also would need to generate their subject knowledge.

The second part of the problem is how it is disseminated and what it looks like for the future. Teacher recruitment is not doing very well. ITT—initial teacher training—recruitment is not doing very well. If we end up with teachers who are teaching non-specialist subjects, it makes situations like that more able to happen.

Tracey Crouch: Matthew, you wanted to come in.

Dr Glanville: That is absolutely fine.

The point I wanted to come in on is right back to your initial premise that the two of you had produced a piece of work of the same standard. Part of my job and all the checks and balances I have put in place are to make sure that you receive the same credit for the quality of that answer. You would not get a lower mark than your colleague because of teacher expectation or anything like that. We would need to do everything to make sure that we had the same issue.

What it might throw up is that, when I am looking for anomalies that highlight things that look out of place, a piece of work where you have done much better than you usually do is probably something that justifies greater investigation because the outcome of you using ChatGPT has been something outside the expectations, and then we have to be very careful to make sure that we do not oversell that. That is why I think it is so difficult to spot where ChatGPT is going because simply the fact that you have done really well is not an indication of you doing anything other than working very hard on this piece of work.

I need to be confident that the level of expertise comes from talking to a teacher who has taught you and can then do that great investigation. There is a potential challenge here to say that you will get challenged on it more likely because it looks out of order, but not that you would get a different outcome.



Q323 **Tracey Crouch:** Joel, you clearly embraced AI particularly in terms of processes and everything else. Do you think there are any aspects of teaching where AI should not be used?

Joel Kenyon: Anything pastoral. When I was looking through where you should not put AI, pastoral was just an instant one—the care that we give to pupils and the small conversations that we have with them when we build those relationships. It just should not come anywhere close. If a pupil has come to a head of year—I am thinking in the dystopian—with an issue and they go on and type in their issue, and get how to fix that issue and advice for it, they are missing the personal connection, and the pupils therefore do not make social connections or learn empathy. They need to learn all those things through a human. There are no two ways about it.

You cannot learn empathy from a computer. If you were to regulate anything, it should not touch pastoral because it needs to highlight the importance of the human connection between a teacher or the support staff and the student. If you were to introduce AI into that, you end up in this dystopian future with pupils who are cold with no empathy and just mimic a computer. Of all the different aspects, pastoral needs to be the one that we make sure AI does not touch.

Tracey Crouch: I presume you agree with that.

Dr Glanville: I absolutely agree with that. I thought the previous point about levels of risk is really challenging. Where you have a huge impact on a student with a decision, that is where I start being very nervous about using artificial intelligence to make that judgment. In the case of academic integrity, it may be a wonderful tool to get evidence, but it is probably not the tool that we would want to use to make that final judgment.

I also thought Rose's previous comment about the difference between teaching and tutoring was a really valuable one, because at the moment there is a lot more that a teacher can bring than just the evidence that artificial intelligence can generate for the teacher.

Q324 **Dawn Butler:** Thank you for my As, Tracey.

To pick up on something Joel said in regard to children mimicking computers, I read that primary school teachers were having problems with children just telling them to do things because they were used to talking to their Alexa, so there was no "please" or "thank you". It was very straight and quite harsh, so they had to unlearn that if they wanted something done.

Do you think that we need to focus on teaching a lot more critical thinking in schools now? If young people are getting information from AI, in order for them to understand that that might not be correct they are going to need to learn critical thinking skills. Do you agree, Matthew?



Dr Glanville: Absolutely. This is not a change in the IB philosophy. We have always focused on teaching those critical thinking skills through our theory and knowledge course. It is critical thinking together with recognising that there are different forms of the truth and why different people hold it, and all of that stuff. It is not new. It has always been important, but with artificial intelligence it becomes even more important and even more urgent that people understand that.

Joel Kenyon: I have two different perspectives. One is I do not believe you can just teach critical thinking. You cannot make a curriculum on critical thinking. That is near on impossible to do. We have our ethical questions in science. We have discussions to do with evolution and creationism, big bang theory versus steady state theory, and all these different types of conversations that make them develop critical thinking. Maybe an emphasis could be made on how that critical thinking comes out in different topics, but I do not necessarily believe that you can teach critical thinking in a curriculum as something concrete.

It forms a part of the curriculum now, and we as teachers do teach critical thinking. Could it be done more? Yes, I believe it could. Should AI now be brought into it? Let us say an AI is making a decision. Should that decision be taken by the AI or a human? Yes, these are now new conversations that we can have. Critical thinking is taught. We teach it all the time. Maybe we should place more emphasis on how the critical thinking is done and why we are thinking in those ways.

Q325 **Dawn Butler:** That is also going to help people to identify the biases that are in the system.

Is there a difference between your vision of AI in education in 10 years and what you think the reality might be? What would you like in regard to AI in the next 10 years, and what do you think will happen?

Joel Kenyon: What I would like to happen is for the situation to remain quite similar, but it just becomes a new tool that we can use, especially to do with feedback. If I have a classroom of 26, 27 or 30 pupils, or in some extreme cases—not in my school but in certain schools around the country—where they are pushing 30 to 33, it is impossible to give the level of feedback that we would like.

In my ideal scenario, when a pupil has answered a set of questions they could receive feedback on how to improve based on the way they have answered those questions, or maybe with their tick-box questions it tells them a topic that they should revise. Ideally, that is what I think it should go to.

In reality, it would appear as a fad that people then start using, a few things appear, and then the ones that stick stay and the rest of them just fizzle away. In 10 years' time, there will be programmes that teachers can use, but they will not necessarily be the best programmes that we want.



HOUSE OF COMMONS

It should be something that is made in collaboration with Government, teachers and educators. What do we want to get out of it? We want to reduce teacher workload. We want to make sure pupils get good feedback. We want to make sure that assessment works. If we left it open in a vacuum, I do not think that would happen at all. It needs to be curated by teachers and Government.

Dawn Butler: The reverse of what is happening now.

Joel Kenyon: If you were to go on Twitter, you would think that AI was taking over. I do not necessarily believe it is. I just think it is being used more as a tool, and we need to work with those tools. We could end up in a dystopia. I do not think that is happening. No one wants the dystopia to happen.

I am also an AQA examiner. We do not look at the tick boxes. The computer does that for us. That is quite a basic AI. That is a reducing of workload and making processes easier. We are already using it. We just need to use that type of thing more. I do not necessarily think it is the opposite of what is happening now because it is happening now. It is used when marking exams. It is used in giving feedback for certain websites. It should be brought into the classroom more often.

Dr Glanville: I think artificial intelligence will be used more and more behind the scenes of organisations like my own with the other exam boards to do a lot of good stuff there, whether it is checking or understanding how things fit together.

My fear is that we will not see the changes of society properly reflected in where we teach our students—a dystopian future where people do not critically think about statements enough. We have already seen that in some of the social media posts. We are already seeing that in some of the fake news debates.

Unless artificial intelligence is trained on the right information source—and who am I to define what that is?—we are going to find that snowballing. I fear that if we do not get ahead of the curve with teaching students how to use this software appropriately, we may suddenly find ourselves in a very dangerous position.

It is exactly back to your first point: understand the bias, understand the critical thinking and understand the critical reading. That is what we need. It is not critical thinking by itself; you need to have a knowledge base to be able to learn how to think critically. In many ways, it is knowledge base independence. It does not matter if you learn it in this subject or that subject. You need to be able to take that across all the subjects, and that is what makes a really good education system. It is not about teaching people to be great biologists; it is about teaching people to be really great learners and thinkers, and that is where we need to get to.



Joel Kenyon: What is important to think of is: what do you want your education system to be? Do you want your education system to be, “You pass this exam, you do well,” or is your education one that benefits the whole of society? The reason we have our education system now is to level the playing field so that every single person has access to the future that they want. In the past, it was limited. It was based on the money that you had or the connections that you had.

The reason we have our education system now is we give every single pupil in the entire country the same resources, the same knowledge and the same information in order for them to do whatever it is that they want in the future. If we make it where education is, “You learn this for X,” we severely limit the impact that education can have on wider society. Education is for everyone, and it is to make sure that we have an educated population that can make the right decisions and that can put people into power who make the right decisions for them. That is what education is for. It is not a tick-box exercise. It is about building a bigger picture. We talk about critical thinking and critical learning. That type of thing builds a good society, and that is what we need to make sure that education focuses on.

Q326 **Dawn Butler:** I agree with that, although there is a lot of teaching to the test that goes on in schools.

If you can make one suggestion to Government as they consider the future of AI, in education specifically, what would it be?

Dr Glanville: I cannot resist responding to your previous comment. I am sorry. Teaching to the test does happen. An obligation on me as an assessment person is to make sure that the test is assessing what is really important. By teaching to it, we are also teaching students what is really important. A backwash effect from the assessment is so important.

That is what I would also say to all education organisations and people thinking about education. Let us make sure we are assessing what is important. Perhaps artificial intelligence will change a little bit—not categorically—what we need students to have coming out of education. Let us assess that.

Joel Kenyon: If I was to say anything to the Government, it is: try to think of a way of implementing AI that does not necessarily have a detrimental effect, and publish guidance and advice on how schools, organisations and companies outside education can use AI for the better by speaking with people in the field and by speaking with experts on AI. Publish advice and guidance on how to use it in the most effective way and keep it out of certain areas. Do not use it in X, Y or Z—for example, with pastoral.

Dawn Butler: Thank you both very much. Thank you, Chair.

Chair: Thank you both very much for giving evidence today. It has been fantastic to hear from practitioners. You have really helped us, as did the



HOUSE OF COMMONS

previous witnesses, in developing an understanding of some of the complex implications of AI for education. We are very grateful for that. We will return you to your classroom and to your students, Mr Kenyon.

Joel Kenyon: I am sure they miss me.

Dawn Butler: At least they get to see you on telly.

Joel Kenyon: Yes, they will.