Justice and Home Affairs Committee

Corrected oral evidence: New technologies and the application of the law

Tuesday 20 July 2021

10.35 am

Watch the meeting

Members present: Baroness Hamwee (The Chair); Lord Blunkett; Baroness Chakrabarti; Lord Dholakia; Baroness Hallett; Lord Hunt of Wirral; Baroness Pidding; Lord Ricketts; Baroness Primarolo; Baroness Sanderson of Welton; Baroness Shackleton of Belgravia.

Evidence Session No. 2 | Virtual Proceeding | Questions 25 - 38

Witnesses

I: Dr David Leslie, Ethics Theme Lead, Alan Turing Institute; Professor Michael Wooldridge, Head of Department of Computer Science, University of Oxford.

USE OF THE TRANSCRIPT

1. This is a corrected transcript of evidence taken in public and webcast on www.parliamentlive.tv.

Examination of witnesses

Dr David Leslie and Professor Michael Wooldridge.

Q25 **The Chair:** Good morning, everyone. Good morning particularly to our witnesses, whom I will introduce in a moment, but I will start as usual with some housekeeping matters. We have apologies from Baroness Kennedy. The session is being broadcast and there will be a transcript taken, which our witnesses will have an opportunity to review and follow up in writing. In the usual way, I would be grateful if members could keep themselves muted if they are not speaking. If they want to come in, raise hands and I will call on people other than those whom I know are asking specific questions, if I can, depending on the timing. I want to finish this meeting by midday, so that gives us five or six minutes for each question and then about 15 minutes at the end.

I welcome Dr David Leslie, who is the ethics team lead at the Alan Turing Institute, and Professor Michael Wooldridge, whose title takes up two lines: head of department of computer science and professor of computer science at the University of Oxford. Welcome, both of you. We are grateful to you for helping us.

I will start with the first question. To give you an idea of the appallingly ill-educated position from which the Chair at any rate comes, I will say that when I buy a lawnmower—which I have not, but as an example—I wonder why I then get bombarded with adverts for further lawnmowers. It would make some sense to me if I was then offered a good deal on a hedge cutter, or weedkiller or whatever. That is my unadvanced state. We want to look at ways of working with advanced algorithms. Can you start by explaining what an algorithm is? Professor Wooldridge, you have leant towards the screen, so let me come to you first.

Professor Michael Wooldridge: Computers are just machines for following instructions. That is absolutely all computers do. They can follow instructions incredibly quickly, but the instructions that you give them have to be incredibly finely detailed instructions, very precise instructions. So a list of instructions to achieve a particular task or to carry out a particular task is all an algorithm is. It is a very finely specified list of instructions, a recipe that a machine can follow. An algorithm is just a recipe for solving a particular problem. A computer program is just an algorithm that has been written down into a particular form that a computer can understand, using a special computer language.

Everything we do, what we are doing now on Zoom, with the email, the web browsing, and so on, all of those computer programs, all of those apps, are just very long, very finely detailed lists of instructions to do certain things. That is basically all an algorithm is, a recipe for doing something, a very finely detailed recipe.

The Chair: Dr Leslie, do you want to add to that?

Dr David Leslie: The only thing that I will add is that we can think of these sets of rules as those that are guiding a process of problem solving. Typically, it is a human or a computer that is carrying out a procedure for mapping a set of inputs to a set of

outputs. Computers typically used to carry out complex algorithms by executing sequences of instructions, but a human could also follow an algorithmic process, such as when someone follows a recipe, where inputs would be the uncooked ingredients and outputs would be the concluded components of a meal.

It might be nice to think historically here—especially as I come from the Alan Turing Institute—that it was Turing who first clarified this, and in the process initiated computers and the digital age. In 1936 Turing had figured out the solution to the problem that had stumped a generation of mathematicians, which was the question of how you define an effective calculation—the question of how we define an algorithm. Turing did this by using a very simple thought experiment. Think of an image of a linear tape divided evenly into squares, a simple list of symbols and a few basic rules. Using this, he drew out a sketch of the step-by step process of how humans carry out any sort of calculation.

When we think of the word "computer", it originally meant a human computer, some individual human being who was maybe carrying out arithmetic. Turing showed that that kind of simple sequence of symbols and using instructions can explain the simplest operations of arithmetic to the most complex equations and how we carry those out. In a sense, what I am describing here is what is known as the "Turing machine". This simple structure of the algorithm showed that how computers ultimately compute is analogous to how we would carry out a computation, and in a very generalised way. That is the source of how we understand the computer age. It is just the simple application of symbols to problems.

Q26 **Baroness Pidding:** Thank you for that. I come from a very similar place as the Chair in my lack of knowledge in this area. Can you explain how advanced algorithms work and what is special about artificial intelligence? How would you define artificial intelligence?

Professor Michael Wooldridge: It is a very big question: how do you define artificial intelligence? We argue about it frequently. The first thing to say is that there are many different definitions of artificial intelligence. In computing in general we try to get computers that can do more and more of the things that human beings can do. Some things, such as the example that David just gave of arithmetic, are very easy to get computers to do, because back at school you were taught the recipes for arithmetic, the instructions that you had to follow. Some of them are easier than others, but nevertheless they were just a list of instructions and you did not need to be smart to be able to follow those instructions. If you followed those instructions precisely you would be able to do arithmetic. But some things paradoxically that we find very easy as human beings turned out to be incredibly hard to turn into the recipes, the list of instructions, that you need to give machines. It turned out to be really hard to get computers to do them.

The classic example that I like is driverless cars. At some point in the future, driverless car technology will really work and will be commonplace. So why is it so hard to get computers to drive cars? It is not the Highway Code, because that is just

a recipe, a list of instructions. Knowing the rules about when you speed up or slow down or signal left and so on is the easy part and we can write down the instructions for that. The hard part turns out to be knowing where you are and knowing what is around you: knowing that the thing in front of you is a pedestrian, that that is somebody on a bicycle, that that is a dog that has just run out into the road—understanding what is around you. We all take that for granted by and large; most people take that for granted. But that problem of perception turned out to be by far the biggest challenge, because we do not know what that recipe would look like: the list of very detailed instructions to interpret what is going on around us.

On advanced algorithms and AI, I said that in algorithms we give a very detailed list of instructions to the machine and the computer just follows those instructions. The different thing about AI, what the current technology does—this is something that has really taken off, it is an old idea that has taken off in the last 10 years or so—is that instead of giving a detailed list of instructions you show the machine what you want it to do. You train it. You just give it lots of examples, "In this situation, this is what I would do. In this situation, this is what I would do". The machine learns from what we show it about, "In this situation this is what I would do, in this situation this is what I would do".

A simple example of that, to make it concrete, is recognising faces in a picture. If you use social media and you upload pictures to social media, and the social media can recognise the face in the picture—you or your child or your sibling or whatever—the way that that works is that you have trained the machine, you have uploaded a picture of yourself or your sibling and you have told it the name of the person in that picture. You do that repeatedly and the AI algorithms on the social media learn to recognise that face. When I see that face I produce the text "Michael Wooldridge": I label it with that text, "Michael Wooldridge", identifying the individual.

So the big difference between the advanced algorithmics and AI that is happening right now, and is getting a lot of people very exercised, is that instead of giving these precisely detailed lists of instructions we are just showing the machine what we want it to do and it is learning how to do it. That is the big difference.

Dr David Leslie: I am almost hesitating about wading into the definition problem, but we need to do it. Again, let me just quickly start from an historical point of view to contextualise the question. If you think back to 1956, which was this famous year where there is the so-called Dartmouth convention where we could say the term "artificial intelligence" was coined, there was no general agreement at that point about what we should call this complex algorithmic processing, or however we want to think about it.

At that time there was a conflict: those who came from more of a cybernetics point of view, who had been influenced by thinking about system-level problems, wanted to call it "complex information processing". We would not call it "artificial intelligence", we would call it "complex information processing" and that would be

closer to the native character of what was going on. On the other side of it, there were some other thinkers at the time who liked the term "artificial intelligence" because it had a punch—we could call it a kind of marketing appeal. The notion that intelligence could be reproduced in machines was appealing to some. At the time, the information theorist, Claude Shannon, who was in this circle—I believe my memory serves right—was hesitant about this move to talk about intelligence in a scientific venue. If you think about it, the word "intelligence" is not what we would conventionally call perspicuous language. It is a complex term that could have many different definitions, very easy to reify or to treat in ways that might not be falsifiable or reflective of the scientific method. Think about this problem. What was picked up in history was the marketable name "artificial intelligence".

That being said, I would step back, ask us to step back and think about how we define AI. There are three ways that we define AI in the contemporary world. These are correlated to three questions: using the question, "What is AI?" as a source of defining it; using the question, "What do AI systems do?", so providing a definition that talks about the function of AI systems; and using the question, "How are AI systems possible?", so thinking about the definition of AI in terms of what are the enabling conditions. To quickly go through those—

The Chair: We do need to keep going.

Dr David Leslie: All right. I will quickly say that if we think about those three routes, first, the "What is AI?" question forces us to define what intelligence is. We might call this, in a fancy way, a kind of ontological definition, a top-down definition that specifies what AI is. It could be rationality, it could be goal-producing behaviour in an environment, or whatever. The second definition asks us to think about what AI does. Artificial intelligence—we might define it this way—is the science of making computers do things, function in the world in ways that require intelligence when done by humans. That functional definition is one that has gone a long way. The last one, in philosophy we would call it "modal definition", is basically a bottom-up definition: what are the basic elements or ingredients that make the design and use of AI systems possible and what are the components necessary for AI systems?

Here we could say, listing them, there are data resources or inputs, algorithmic or computational methods, processing delivery and deployment platforms, knowledge, practices, problems, spaces. All of those things are important for us to think about when we are defining what artificial intelligence is, because all of those things are necessary components of how AI works in the world. Those three angles are important.

The Chair: Baroness Pidding, shall I go straight on to the next question?

Baroness Pidding: Yes.

Q27 **Baroness Hallett:** What are the requirements for an advanced algorithm to achieve its intended purpose both consistently and reliably?

The Chair: Shall I start with Dr Leslie this time? As I say to both of you, I am sorry, these things always deserve about four times as much time as we have available.

Dr David Leslie: Sure. I will run through a few quasi-technical dimensions that we would associate with an algorithm being able to achieve intended purpose. Let us call these accuracy and performance, reliability, security and robustness.

Accuracy and performance, we understand as a model of success, the proportion of examples for which it generates a correct output. For a consistently well-operating algorithm to be that, it needs to have a high level of accuracy and high level of performance. We could go through other different performance metrics, but that dimension of accuracy or performance metrics will always go into the way we measure success or reliability.

To delve into the reliability dimension, we can think about reliability as an AI system that behaves exactly as its designers intended or anticipated. A reliable system adheres to the specifications it was programmed to carry out at design time. Reliability is a measure of consistency and can establish trust or confidence in the safety of a system, based on the dependability with which it operationally conforms to its intended functionality.

Security is a third dimension of this, which encompasses the protection of several things. A secure system is capable of maintaining the integrity of the information that constitutes it. This includes protecting its architecture from unauthorised modification or damage. A secure system also remains continuously functional and accessible to authorised users and keeps confidential and private information secure, even under adversarial conditions.

Finally, robustness—and this is a really important one. The objective of robustness can be thought about as the goal that an AI system will function reliably and accurately under harsh conditions, unexpected conditions. These conditions can include adversarial interventions, implementer error. If you are thinking about what are called “reinforcement learning applications”, there could be something called “skewed goal execution” where the system is not quite following what the objectives of the designers might have been. Measures of robustness are the measures of the strengths of the system’s integrity and soundness in response to difficult conditions.

All of those components—robustness, security, reliability and accuracy—figure into the way we would think of a technically well-performing algorithm. I do not know if I hit it, but Michael will cover what I did not.

Professor Michael Wooldridge: Let me focus on one very specific, very narrow but very important aspect of the question, and that is to do with the advanced algorithmics, the AI-type stuff. There are two things that I want to try to highlight.

First, I talked about the way that with this new technology you train the computer to be able to do something. You give it lots of pictures of Michael Wooldridge labelled with the text “Michael Wooldridge” and eventually the hope is that you will be able to show it a picture of me and it produces the right output. The first thing to say here is that the technology is very brittle, and what I mean by that is that it can

fail in bizarre ways. You can make a tiny change to a picture of me that we would not even notice and then all of a sudden it labels that picture with "Lord Blunkett". That might not seem so bad if you are just on social media and it is mislabelling something but—going back to driverless cars again—it turns out that you can make very subtle changes to road signs and the very clever advanced algorithmics in the car, which is all to do with reading the road signs and understanding what they are, gets it completely wrong and it fails totally. So the technology is very brittle.

It is worse than that, it is brittle in ways that we do not understand. We just literally do not understand what makes it go wrong. And we cannot predict when it will go wrong. Does that matter? If it is just mislabelling me on social media; no, it does not matter. An example of a successful advanced algorithmic technology is automated translation. Take a text in Chinese and translate it into English—and this technology works quite well. The last time I was in China the air conditioning in my hotel room was set up way too high and the instructions were in Chinese. I had an app on my phone that translated, not brilliantly but well enough so that I could control the air conditioning. Given what I have told you about how brittle the technology is, would you trust it to translate an instruction manual for a Boeing 747? No, that would be very foolish, of course you would not do that. The technology is very brittle.

Again, it is brittle in unpredictable ways. But it is even worse than that, because we do not understand how it works. With the current AI technologies, the way that they are built is that they are reduced to very long and very opaque lists of numbers. The only technical thing I will say is that they are what is called "weights" in a neural network. Ultimately, it is just a big long list of numbers, and that list of numbers does not mean anything to anybody. We do not know how to interpret that, we cannot extract the knowledge that this system has, the expertise it has, we cannot query it in any meaningful way.

So the technology is brittle, it is brittle in unpredictable ways, and this technology is a black box. We do not really understand how it works. We just hope that when we give it the input—the picture of me—it produces the right output, or that if we give it the Boeing 747 instruction manual, it translates it correctly.

The Chair: Baroness Hallett, do you want to—?

Baroness Hallett: No, I am terrified.

The Chair: Me too.

Q28 **Lord Dholakia:** How can humans make best use of artificial intelligence? What should advanced algorithms be used for and what should they not be used for? You mentioned just now about the technology being very brittle. Should there be a code of practice about the use of artificial intelligence?

The Chair: I will suggest to the two of you, and it will be frustrating, that the one who does not come first just fills in any gaps that you think there might be.

Dr David Leslie: Sure. It is a great and large question. First, if we think about the warning that Mike just gave, that is a mixed bag in the sense that the warning is also

the source of some of the advantages of the technology, which is to say that, because they are so complex—we say "high dimensional" because there are so many variables that can get thrown in there—advanced algorithmic systems can identify complex patterns that simply escape human-scale comprehension.

Humans can hold three to eight variables together in our head at any given time and think through the relationship of these variables. We can maybe think in four dimensions. We have our three-dimensional spatial orientation and we have the fourth dimension of our temporal orientation. Advanced AI systems can think in almost unlimited numbers of dimensions and can concatenate and put together tens of thousands, millions, billions of variables at a time.

That means that certain of these systems can pick up patterns that would simply escape the kind of way in which we understand the world. It is not to say that that will always yield positive, but if we are thinking on an epidemiological scale or thinking on a cosmic scale, where you have a massive amount of data that can be processed, an algorithmic system that is processing in a high-dimensional feature space with multiple variables can support new insights by generating outputs that we would not have come to, in that we would not have been able to pick up on the patterns. That is something to keep in mind, with measure, because if the system, as Mike was saying, is not explainable or interpretable, we also face problems of opacity that mean we need to think about how we understand the patterns that are being generated.

Also, this signals a warning about good and proper use and improper use. Say our complex system is not processing astrophysical data or biochemical data but it is processing social and demographic data—data from what we do in the world, how we move around in the world, what our characteristics are. If you have a complex system of an order that is making correlations that we do not understand, that system can reproduce discriminatory patterns, bias, in the world without us ever being able to detect it.

So in certain instances we really need to be careful. When we apply complex algorithms to our social world, we need to have interpretable systems, because how else would we know if the system is being properly non-discriminatory or fair? We need to be careful that in putting algorithms to use we are aware of the potential for lurking variables, for the creep of harmful patterns into the system. We should not necessarily use such complex systems when they might preclude us from identifying those kinds of harmful patterns.

There are a lot of advantages, but I will leave it to Mike because I do not want to talk too long.

The Chair: I was going to ask Lord Dholakia if he had any supplemental arising from that, which I can then put to Professor Wooldridge, or shall I head straight to the Professor?

Lord Dholakia: I am happy so far with the answer.

Professor Michael Wooldridge: Very briefly, where particular care is needed is where AI is making decisions that have substantial consequences for human beings. David is right to highlight that the naive use of this technology can lead to systems that are inherently biased. We have talked about training these machines. If the data that is being used to train these machines includes human biases, the machine will reflect those biases. There is now a very depressing and very long history of evidence where that has happened.

The EU has some proposals for AI regulation. They are not perfect but they are an interesting discussion point. They identify a number of what are called high-risk areas. I will not go into those, but they include things such as facial recognition and so on. You might want to take a look at what they identify as being the high-risk areas as a useful starting point.

I will give you one example, just to conclude. The police force in Durham started introducing a system called HART, which I believe stands for "harm assessment reduction tool". Essentially it advises custody officers on whether somebody should be kept in custody in a police cell or released. What they did is use classic advanced algorithms, AI machine learning. They took custody records going back a number of years where somebody had been released and it was a good outcome, no further crime was committed, or somebody was released and it was a bad outcome, and they used that to train a machine-learning program. The idea is that when somebody is in custody you can use all the data, whether there is a history of offending, whether they are a substance abuser, whatever, and the machine makes the prediction.

I have looked at that system quite carefully and it is a very thoughtful and careful piece of work; there was a lot of diligence in that system. But my worry is—although it is clear that this is supposed to be just another voice in the room, just making another voice to be heard—that that becomes eroded over time and that in the future you have tired or confused custody officers who are facing a difficult situation, and the machine says, "Keep them in custody" and you abdicate your responsibility and go straight to making those decisions. So I think that overreliance on the technology, for the reasons that we have touched on, is a real area of concern. People need to understand the limitations of it and need not to over-humanise the technology. It is just at best another voice in the room.

The Chair: We have heard the term "human override", which I think we will do well to bear in mind. Baroness Hallett has her hand up.

Q29 **Baroness Hallett:** I have just one question. Professor Wooldridge, you talked about a number of biases that have been identified. Could you give us an indication, very briefly, as to the nature of the kind of biases that have been identified?

Professor Michael Wooldridge: There is a great book on this subject called *Invisible Women* by Caroline Criado Perez—I am not sure how to pronounce her name. Here is an example. There are standard what we call "datasets"—training data that has been made available over the years for training your program. Imagine that you are

trying to train your computer program just to label pictures, the kind of pictures that you might take, and your training data contains lots of pictures of kitchens, which is a standard example. If it happens that lots of the pictures of kitchens contain women, your program will learn an association between women and kitchens and, if asked to produce a picture of a woman in the future, the program might well produce a picture of a woman in a kitchen. This is a real example.

Another brief example of the training data problem is the datasets that are used to train speech recognition: that is, being able to listen to speech and interpret it. AI researchers historically have been, by and large, white, college-educated males, very often American, and the training data was produced by these people. So it consisted of training data from white, male, college-educated Americans and when it is given voices that fall outside that rather narrow demographic, the program has not been changed to recognise those. So women and ethnic minorities end up with the technology not being able to interpret their voice. There is a huge number of examples of that, and the book that I mentioned is a very good introduction.

Q30 **Lord Blunkett:** From your comments about training to be able to establish and use the machine to education and training for the public, in the last half hour you have been doing to us what my question is all about. How on earth are we going to establish, and are there any very good examples already established, education and training programmes—you have mentioned the European checklist—to help people to use this technology? My guide dog is a very basic learning machine in the sense that it got trained and then is rewarded for using its initiative—but I would never allow it to attempt to cross a road unless I was engaged with it as well. That might not be a good example for the wider public, but it certainly enables me to see over the last 35 minutes what you are talking about. Are there good examples of education and training that we could draw down on? We will be talking not about coding and how we develop the capacity of the public to be able to contribute to this, but how they understand both the possibilities and the challenges. For instance, in the Stock Exchange there have been occasions when an algorithm has virtually brought down the international money markets.

Professor Michael Wooldridge: There are a number of public education programmes. Like David, I am also involved in the Alan Turing Institute and it does a number of these public engagement activities, and we are busy ramping up our programmes. I think we are still finding our way with the technology, because the technology is evolving as quickly as our understanding of the issues that we have been talking about. I would be very happy to provide the committee, offline, with some guidance, and I am sure David would be as well, on places to find out about these things.

But if I had to boil my advice down to what is going on with this technology and how to interpret it and interact with it, number one is, “Do not humanise it”. It is not like human intelligence. These are just programs that have been very finely tuned to do one tiny little thing very well, like recognise a face in a picture. They have a very narrow focus. They do not have any understanding of what they are doing. Do not

imagine for one second that there is any comprehension of the task that they are trying to do. Do not humanise it; there is nothing like human intelligence there at all. “Forget that completely” is my advice. Be sceptical about it, query it. Do not just accept the recommendations, like the Durham police force example that I just used. Do not accept those recommendations. And this is quite a tough thing for a lot of people to do. It is the famous “Computer says yes, computer says no”. We are very used to just listening to it and we need to learn how to say no to the computer. So do not anthropomorphise it—it is just a very finely tuned computer program—and learn how to say no to it. Be sceptical, question it.

Dr David Leslie: Perhaps I can fill in a couple of things. I would enjoin you to think about the training and upskilling as involving both a technical and an ethical or moral component, which is to say that we do need technical upskilling when it comes to training implementers and users and affected decision subjects—data subjects—to understand the limitations of statistic or probabilistic systems. These are systems that are built on uncertainties and, if we do not understand the limitations, we will tend to either over-rely or over-comply, which are implementation biases, or we can even disregard the results when we do not fully understand the strengths of a system. So we need to have the technical upskilling of those who will use the systems and be affected by them. On the other side we need a moral or ethical upskilling, so that those who are involved in designing and developing these systems see the processes of building them as sociotechnical processes. Practices where there are ethical and moral consequences of behaviour in a social environment, the choices of how we might tune a model, are choices that have direct ethical and social consequences. Choices of how we formulate the problem that a system is solving have direct social and ethical consequences. So we need to also develop that dimension of training and upskilling.

I will quickly mention that I think that the UK is in a particularly advanced position here, because right now the national data strategy is already trying to institutionalise modes of training. I know that at the Alan Turing Institute we have started to beef up our development of curricula that we can use for civil servants and early career researchers to steward this process of technical and ethical upskilling.

I will say one quick last thing. If we wanted to look at good examples of how upskilling can really be a benefit, about two years ago the ICO and the Alan Turing Institute held basically citizens juries to inform the construction of our guidance on explainability in Manchester and Coventry. I oversaw the process and, to the person, those who had been involved in the two citizens juries stressed the transformative effects of learning about the basics of statistics as a way that they could then make informed decisions about what they expected from explanations of these systems. So it is really important to keep in mind that there is a cognitive equity dimension here where we need to be training citizens, not assuming that they will just not be understanding. We need to assume that always we can go to the lowest common denominator and make it clear that this is just mathematics at the end of the day.

Lord Blunkett: I think what has been offered is very helpful. If offline we could have further information, I am sure we will want to come back to this later in the inquiry.

The Chair: Yes, I am sure we will. I will bring in Baroness Shackleton, although I think largely the question has been answered.

Q31 **Baroness Shackleton of Belgravia:** I think the question has absolutely been answered. The question was whether AI performs better than human intelligence but, from what Professor Wooldridge is saying, it is in conjunction with and married together. There is an obvious example of that. If you have the ingredients of eggs and chocolate and flour and butter and you shove it into a machine, you have to tell it whether to make a souffle, a cake, little buns. Without somebody telling it what to do, you will get a mixture that is not fit for purpose, without the human being actually doing it.

I am more interested in Dr Leslie was saying about morals. We are all educated, hopefully, to know the difference between right and wrong. That is not always easy to apply, because it depends on judgment. When you are the car going down the motorway without a driver and there is something that looks like a bull—nobody knows what it is, it could be a bomb, it could be a balloon—somebody has to judge whether it is worth swerving the car to the right or to go through it. There is no absolute right. How are you ever going to be able to train the artificial intelligence to have morals to make an informed decision on either the best or the least-awful decision to take at very short notice?

Dr David Leslie: These dilemmas have had a lot of play over the past couple of years, especially thinking about how an autonomous vehicle would handle that type of problem. The way I look at it is that what we need to focus on more right now is not questions about the point of decision with automated ethics determination, but how we can allow for the designers and implementers of systems to have ways of adjudicating between conflicting values—values that are coming into tension.

In this instance we need to bring into the processes of designing, developing and using systems procedures for having open and mutually respectful conversations about how our values will inform the direction of travel of the technologies themselves. We have a lot of different values. I might value human dignity or social connection or public security or the priority of social justice and you might have a different view of how we should weigh those. But what we really need as we set up designers, developers and users to make human values operational within the context of the systems is procedures for coming to conversational agreement about how those values should be put into the system. That involves having a common starting point in an ethical vocabulary. This is why you have seen so many different ethics frameworks of five to 10 to 15 principles, because we need a starting point to have those ethical conversations—but we also need the proper procedures to think through how those values and ethical principles might relate to one another.

I do not think there is a simple answer to the question, but I can say that in modern life we move around the world where we are not able to appeal to the absolute cosmic order or the dictates of religious doctrine when we are trying to navigate the

world together. We have come up with ways that we can have meaningful conversations about reasonable action. We need simply to put those mechanisms into practice when it comes to the ways in which the systems are designed and used. I skirted the question a little, but I think that is what we should focus on right now.

Professor Michael Wooldridge: I will give a slightly different perspective. There is a very standard idea in philosophy, the notion of a moral agent. A moral agent is just an individual who can distinguish right and wrong and understand the consequences of their actions with respect to right and wrong. I do not think we should think about AI systems as being moral agents; I do not think that is somewhere we want to go. The reason for that is it will allow people and organisations and Governments to abdicate their ethical responsibilities. I do not want computer programs to be ethical, I want the organisations that write and deploy those programs to be ethical. I do not want to get to a place where there is a loophole where a Government can say, "Our drone was very naughty. We trained it to do the right thing and it somehow made the wrong choice". No, that is just abdicating your moral responsibilities.

I do not think it is at all realistic any time soon to think of AI systems that could conceivably be moral agents in that sense—but, in any case, I do not think that is where we want to go. We should view these things as tools and it is we who are responsible for the ethical aspects of those systems.

To go back to the driverless car, it is a famous example that we have seen many times. What is actually going to happen is that the code that the engineers very carefully write is designed to be risk averse and to minimise the risk. It will try to do the safest thing possible and if it encounters a situation that it does not understand it will just get itself to a safe place. It certainly will not start doing any ethical reasoning—there is a nun on this side of this road or a freckled child on the other. It is just not going to go there and I do not think that we want to go there with that technology.

Baroness Shackleton of Belgravia: So the answer from both of you is that AI is not superior to human intelligence but it is a very useful tool that human intelligence can deploy morally. Rather than creating the wheel every single time, you can look at a set of data and it is for the human brain to analyse how to use it. I think the fear for junior people coming into office is, "Am I going to have a job in a few years' time? Am I going to be replaced by AI? The job will not exist because somebody will be doing it better". The answer is no, because you will always need a human to feed the instruction and to analyse the result.

Professor Michael Wooldridge: Are we going to visit the general issue of employment and things like that at all?

The Chair: No, it is not really within the remit, but I can see how it is an example that makes it real to ask.

Baroness Shackleton of Belgravia: Thank you both very much.

Lord Blunkett: But it is happening in banking where, after the crash in 2008, they are relying very heavily on AI that monitors and predicts and in many ways oversees the behaviour, rather than the behaviour being cultural from the top down. I thought what Professor Wooldridge said is absolutely spot on.

The Chair: It is, and you are right; that is a good example of how one might think about it within the task we are setting ourselves. I think primarily it is helpful as an example, but we will come to that subsequently. I will move on to Baroness Primarolo.

Q32 **Baroness Primarolo:** I am trying to process all the information as we are going along and it has been phenomenal so far. Dr Leslie referred, in answer to Baroness Hallett's question, to using a description of performance and accuracy, reliability, security and then robustness, and defined how he was using robustness. To take this down to a level of "How do we know?", the question is: how can the performance and robustness of the advanced algorithm be assessed? We have talked about all the shortcomings and what might be the potential, but to know the potential we need to have that loop back into: is it doing what we wanted it to do or does it have the unpredictability and the black box that Professor Wooldridge so eloquently explained as well? I am a bit stuck here. It sounds like a good idea, but how do we know, how do we assess?

The Chair: Another problem is that we will have to ask our witnesses to explain this at quite some speed, which will make it harder for us—but let us see what you can do. Dr Leslie first, let us see how quickly you can answer that question.

Dr David Leslie: Really quickly, I think that all of the issues that Mike has been underlining—black box issues, opaqueness issues and complexity issues—are real issues when it comes to thinking about the assurance of these elements of the systems. We know that when you have a system that will be functioning in an open environment and might not have a frozen model, it might be a dynamic learner, that you are not going to be able to prespecify intended functionality. The system might be unpredictable as it relates to the uncertainties in its environment, the unknown unknowns. To compound that, we might be in a situation, as Mike was saying before, where the actual complexity of the system does not yield a viable explanation of why it acted the way it did.

To counter that a little bit, there has been a tradition that we can think of as argument-based assurance, building assurance cases, whereby we think about a set of top-level priorities or normative goals. Say we take it at the top level; we want to assure these qualities of reliability, robustness, safety and security. If we are trying to assure those and we are on a design team we would have to think about what properties we need to put into our system as we build it that would assure that it is safe, secure, reliable and robust. To assure that, to provide some sort of confidence that we have done that, we would need to take and document actions that will assure, say, the robustness of a system. Say I wanted to assure that my system was going to act robustly in uncertain circumstances. I might act to harden my model and that action would be part of an assurance case that would then, hopefully,

provide confidence that the properties of robustness had been taken into account going into production.

I am explaining one way that, by focusing on assuring properties, taking deliberate actions and presenting it as an argument with documented evidence, we can move some way towards building a little bit more public confidence in the use of these systems. That being said, I really stress what Mike said before: that it does not solve the brittleness problem and some of the other uncertainties. But it moves in the direction of having more responsible practices of designing, developing and deploying the systems.

Professor Michael Wooldridge: To very briefly add to that, I would make a distinction with what I call conventional software. The vast bulk of the software that we use—Microsoft Word and PowerPoint and so on—is very conventional, going back to my introductory thing about lists of instructions. We are not anywhere near perfect with that kind of software, but we have lots of experience of how to build it robustly and correctly. An extreme example of that is the airline industry. One of the reasons that the airline industry is so safe is that a lot of what goes on when you fly an aircraft is to do with computer code taking control and it is phenomenally safer now than it was 50 years ago. We have a pretty good handle on that with conventional programs

I emphasise again that we really do not have a handle on the advanced algorithmics, the AI machine learning, the training things. There are many brilliant researchers internationally working on the challenge of dealing with brittleness, the unpredictability and so on. We do not have answers yet and so the very big caution is that where that technology is used, be extremely careful, extremely cautious about using it naively in places where it has consequences for human beings. I guess that is my main point.

Baroness Primarolo: I just want to ask another quick question. If AI is interactive and it learns and if you use whether you call it an argument-based assessment that it is working properly or some sort of stress test, for want of a better word, how do we know that it does not interact with that challenge and, therefore, give a different reaction and go into the black box that magnifies the unpredictability and brittleness of the system? Or is that a daft question? Sorry.

Professor Michael Wooldridge: It is not a daft question. It is a difficult question. It comes down to engineering. These are things where the computer programs, those lists of instructions, are very carefully engineered and it will be down to the software engineers who build the code to ensure that the kind of scenario that you are talking about does not happen. Scenarios where the machines learn in unpredictable ways are now quite carefully studied. One of the difficult bottlenecks is between you and the computer; you want the computer to do something for you but communicating to the computer exactly what you want it to do is very difficult.

Here is an example to illustrate this. I was writing some code. It was very simple code—it sounds quite fancy but it was not—to do with co-ordinating a network of

trains about 15 years ago. It was a really simple, circular network with two trains and there was a bridge on the network. The thing was that if ever the two trains were on the bridge at the same time there would be a crash. We wanted the program to basically learn how to avoid crashes. We did and the way that it learned was just to stop the trains from moving. It came back and said, "Right, here's how you avoid a crash. You just stop the trains from moving". I had forgotten to communicate that that was actually quite an important part of what I wanted the program to do. There are many examples like that and it is a work in progress to be able to deal with that.

The Chair: Thank you. I need to keep an eye on the time, which is not intended to be aggressive towards either our witnesses or members of the committee and certainly not to Lord Hunt who has the next question.

Q33 **Lord Hunt of Wirral:** For many of us the key question in the background of all that Dr Leslie and Professor Wooldridge have been talking about is law enforcement: what is appropriate? What validation and evaluation tools do you think would be appropriate in that law enforcement context?

Professor Michael Wooldridge: I used the example of the Durham HART system earlier. I looked at that quite closely and I think it was a thoughtful and very considered piece of work. The problem is that it is very easy to do similar things that are not nearly so well considered and there are some concerning examples from the United States. There is a program, which I believe is not really an AI program, the gangs matrix program that I believe is used by the Metropolitan Police. The gangs matrix program is profiling people to decide whether they are likely to be members of a criminal gang. The concern with that is whether it should really be making a judgment on whether somebody is likely to be a member of a criminal gang based on the music that they are listening to on their iPod. There are examples where that kind of thing has happened.

So it has a huge role to play but I think what we need to do is very firmly recognise that it should not ever be more than another voice in the room. There is a very big challenge in the way that we deal with this technology that we do not over-rely on it or come to rely on it, either explicitly or just inadvertently because we are tired and so on. That is where the biggest risk is. All the examples that we talked about earlier to do with bias are incredibly relevant here and painfully relevant in these kinds of scenarios in very obvious ways. I think that is probably number one on my list of concerns and many other people's lists of concerns, but there are many other concerns besides. As I say, I think it is recognising the limitations of the technology and not over-relying on it.

Lord Hunt of Wirral: Chair, may I ask Dr Leslie to focus on facial recognition just for a moment and give us his advice? That is described as "the new plutonium" by experts, who recommend that it should be used only when absolutely necessary. We are all talking now about human involvement, to what degree, to determine whether it is appropriate or inappropriate to deploy artificial intelligence. What is your guidance to us on that?

Dr David Leslie: This is a rightfully incendiary topic because, as you point out, the use of this sort of biometric surveillance technology has been called out as being toxic and potentially harmful on many different levels.

First and foremost, we have to acknowledge the legacies of discrimination that have manifested in the type of large-scale datasets that have been needed to train facial recognition systems, where you have, as Joy Buolamwini would say, a prevalence of "pale males", white male figures who have been used to hone the accuracy of these facial recognition systems. You have a starting point that should give us pause, which is to say that these systems perform with differential accuracy with regard to different skin tones and social groups. If anything, that should be a pause button right from the start. If we think we will be justifiably able to use these sorts of systems in the world, we need to be able to guarantee that the way that these things are operating is equitable and non-discriminatory.

That is just the starting point of it. If we then go into the ways in which live facial recognition systems can pose risk to cherished rights to privacy, self-expression, association and consent, we are moving into another area of concern that we also need to pay attention to. Live facial recognition systems may prevent citizens from exercising their freedoms of assembly and association, robbing them of the protection of anonymity, and have a chilling effect on social solidarity and democratic participation. The use of these sorts of systems, AI-enabled biometric surveillance, can likewise strip citizens of their right to informed and explicit consent in the collection of personal data. If you set cameras up in a public space that are recording and then analysing, we are entering into a zone whereby these systems could easily infringe on a wide set of rights and freedoms. We published a paper out of the institute, *Understanding Bias in Facial Recognition Technologies*, and I recommend that. I will include that in what I send you.

Lord Hunt of Wirral: We must look at that. Thank you, Chair.

The Chair: Thank you. Can I move to Baroness Chakrabarti? You have a couple of questions that I think will come together and follow on from that.

Q34 **Baroness Chakrabarti:** Chair, if I may, I might merge them, because we have begun to answer them already.

Gentlemen, you have spoken very eloquently already about how you think things ought to be, avoiding overreliance and so on. Can we move away from how things ought to be to how they are already? You have given the very concerning example just now of facial recognition. Can you give other examples of the relationship between advanced algorithms already operating and our societal norms and ethical principles? How are the algorithms affected by our programming? Equally, perhaps, how are the algorithms affecting our societal norms and behaviours? It may be two-way traffic.

The Chair: And what we should be looking to in the future as well.

Professor Michael Wooldridge: Sorry, I did not prepare my answer to this, so I am struggling to find a good example, but let me give you one. A major technology

company a couple of years ago wrote an AI program that was a chatbot, basically, on Twitter. I think they thought it would be quite fun to have a program that could look at Twitter and learn to communicate. I think the idea was that it would talk about "Love Island", "Eastenders" and things like that and all be a bit of fun. It very quickly turned horrible.

What the chatbot learned from Twitter and the interactions that it had was to be racist and misogynist. This thing was spouting racism and misogyny in the most grotesque way possible within a very short space of time. It was a harmlessly intended experiment that demonstrated how quickly these things can get out of control.

There are lots and lots of fake accounts on social media. We are going to a time when there will be more AI accounts on Twitter than there will be human accounts. How do you deal with social media when everything that is being created there is being created by programs like that? That is one example.

Baroness Chakrabarti: In your example, the bot is now behaving in the bad way that it has learned, but by behaving in that way it is encouraging humans to behave that way in turn.

Professor Michael Wooldridge: Yes, exactly.

Baroness Chakrabarti: It is a vicious circle. Thank you. Dr Leslie?

Dr David Leslie: This is an example of what we might call transformative effects, where the systems themselves are transforming our identities in the process of operating in the world.

I will take a step back for a second and just say that if we think about modern democratic forms of life, we might understand our ability to move around the world as free individuals and as individuals who are participating in the moral life of the community. As modern individuals, we have an element of individuation or becoming autonomous, independent selves, and reproducing the world together through our democratic interactions, through collaboration and through the ways in which we can come together, build consensus and organise our political and social affairs. We are seeing warning signs now that the transformative effects of algorithmic systems are challenging those dimensions of our modern, democratic forms of life.

I can quickly talk about a few examples to illustrate it. Think about the transformative effects already apparent in the broad-scale proliferation of individually targeting algorithmic curation. Incessant AI-enabled personal profiling may generate profits for the enterprise of ad tech, but it also continually produces an impoverished reflexivity and surveillance anxiety that has emerged from this growing ecosystem of distraction technologies and compulsion-forming reputational platforms on social media.

We can also think about the ways in which these kinds of mass-scale behavioural steering mechanisms at the collective level—think here about relevance-ranking, popularity sorting, trend-predicting algorithms that are coming out of large platforms—are producing in a sense calculated digital publics, spaces of digital public that are devoid of active, participatory social or political choice. Rather than being guided by a kind of political will that is discursively and deliberately achieved, citizens are caught up in this vast meshwork of connected digital services that are shaping publics according to this drive to target, capture and harvest individual attention.

There are just two examples of the ways in which individual autonomy and democratic participation are, in a sense, challenged by these at-scale uses of algorithms. I could go on and give others, but that is where I will start or stop.

Baroness Chakrabarti: Forgive me. Are there any more positive examples of a democratic, ethical or governance engagement that has a transformative effect in the opposite direction and chills the more troubling potential of the AI?

Dr David Leslie: Thinking about the examples that I just gave, we can also think about the ways in which digital platforms and algorithmically informed digital platforms have enabled wider and deeper forms of participation. Here I am thinking, for instance, of a platform called Decidim and one called Consult, which are both grass-roots, participatory mechanisms for organising people coming together, giving their opinions, doing collective budgeting for local communities. We can think of instances in which these structures could be used for a positive, democratic purpose, but we need to see ourselves in the driver's seat rather than being subjected to large-scale, extractive mechanisms in the way these platforms are presented to us.

Q35 **Lord Ricketts:** Thank you very much, both witnesses. It is an absolutely fascinating session. Could I ask you, picking up from Baroness Chakrabarti, to look forward to where we are going here? My layman's view is that advanced algorithms have been quite slow to develop so far. They have been talked about since the 1950s, as you were saying, but the ambition has not been matched by usable applications until quite recently. By comparison, say, with computing power and Moore's law that it doubles every 18 months or two years, they seem to have been slow.

Are they now accelerating? Are we now going to see a dramatic, exponential acceleration in the application of advanced algorithms right across the board, including in the area we are interested in of law enforcement? If so, given what you have said about the fact that you do not really understand how advanced algorithms do their calculations and they are extremely hard to audit, what are the implications of that? I am interested in where you think things might be in, say, 10 years' time, if we are in that accelerated phase of application.

Professor Michael Wooldridge: That is a very good question. My personal take on this is that if you go back 40 years, basically none of us had ever encountered a computer and now we encounter dozens of computers every moment of our lives.

What would have been a powerful desktop computer 15 years ago is now on my wrist. Your smartphone that you carry with you in your pocket would have been a supercomputer 30 years ago. My personal view is that basically AI technologies will become embedded in the same way that our world is now populated everywhere with computers.

That will mean that every time you have to make the most mundane decision imaginable, there will be AI there to help you and advise you. You are trying to decide on a restaurant to go in, you look at the restaurant and it reminds you that you went there two years ago and had a bad meal. It gets bad reviews, or it got a bad review or a bad hygiene rating two weeks ago, and it says that there is a better, more well-regarded Italian restaurant 100 yards down the street. Think of AI there as being what I call a cognitive prosthetic. It is augmenting your intelligence and helping you to make better, more efficient decisions. Every place where you make those decisions, it will be embedded.

In the world of work, you will not be interacting with robots; there will not be robot AI systems to interact with. It will just be absolutely everywhere in everything that you do. In all the programs that you use, there will be AI embedded there.

As I say, the key phrase there is cognitive prosthetic. That is how I think AI will manifest itself. The other part is that it will be everywhere. It will be ubiquitous, embedded in everything.

Lord Ricketts: Dr Leslie, in the areas where human judgment has the most impact on other human beings—for example, law enforcement activity or military applications—that feels pretty scary to me. Quite how you would know whether an autonomous weapons system had made a decision that was within the rules of war, for example, I do not know. How you audit these kinds of advanced algorithms in the background of every decision that law enforcement is making feels to me like quite a big problem.

Dr David Leslie: Yes. Standing at the precipice that Mike has been talking about of the coming ubiquity of cyber-physical systems, I would say that we also have to see this as an inflection point as to whether we cede the decision-making capacity to more and more forms of automation. My feeling on where we stand and where we are going is that if we are able to put ourselves back in the driver's seat, directing the trajectories of innovation, we can manage the growing ubiquity of automation in ways that are in accordance with our collectively articulated purposes.

I think it is important for us to realise that we will be able to do that only if we take a principles-based approach where we prioritise democratic practices, individual will and choice. As we move into that future of more and more computational systems all over, we will consciously have to manage the choices, not about rectifying the problems in existing technology but about engaging in anticipatory reflection that can set the direction of travel for the innovation from the start. Rather than reacting to it, it needs to set the direction of travel for the innovation.

As we look forward to that, we can see a path where you have AI supporting, say, sustainable development goals and assisting in many elements of human experience that would lift our possibilities for improving the lot of humanity, but we can also see another track where, if we are not responsible, you have converging issues of algorithmically informed synthetic biology, geoengineering and other large-scale uses of AI-enabled technology that will present global, catastrophic and existential risks. At this point, we should be conscious that we need to put the human being in the driver's seat here.

Lord Ricketts: That is a very good note to end my questioning on. Thank you very much. Back to the Chair.

Q36 **The Chair:** We have set ourselves what is beginning to seem an increasingly difficult task of establishing some guiding principles for the use of technologies in law enforcement. I am using that term very widely. You have mentioned the work that the EU is doing and that it has already published something. A code of practice was mentioned earlier. Do each of you have principles that you would like to recommend, in bullet point form?

Dr David Leslie: For me, it would be a good starting point to take a three-node approach to this. First, we need to start to codify and institutionalise horizontal, cross-sectoral frameworks of principles-based algorithmic governance and documentation beyond the existing data protection regimes, equality law and human rights provisions, because there are additional ethical questions and issues of collective rights involved.

In the form of bullet points for this node, practical principles that set up guard rails to ensure responsible innovation would involve the following: first, socially sustainable AI with designers and implementers being held responsible for their explicit considerations of the transformative and long-term effects of their systems on individuals, society and the environment; secondly, technically sustainable AI where designers and implementers are held responsible for the safety, security, reliability and robustness of the systems they produce or procure; thirdly, accountability and transparency, where accountability entails that designers and implementers are answerable for the parts they play across the entire design and deployment work flow, and transparency entails that the design and implementation processes are justifiable and well documented through and through; fourthly, a fair, non-discriminatory, bias-mitigated and equitable AI where designers and implementers are held accountable for being equitable and not harming anybody through bias.

Fifthly, I would say AI that is built in accordance with high standards of data quality and integrity and responsible data management, following, say, the ONS Five Safes: safe projects—the ethical use of data for the public benefit; safe people—accredited users with proper responsible research and innovation training; safe settings—making sure that access to data is secure and that computer infrastructures are secure; safe data—thinking about responsible data linkage, data integrity and FAIR data, meaning findable, accessible, interoperable and reusable

data; and safe outputs—thinking about controlling the flow of information, sustained and differential privacy, and traceability of data.

Those are the practical principles. The second node is developing vertical and sector-specific guiding principles for the law enforcement sector in the spirit and building of things like Marion Oswald's ALGO-CARE, which I would recommend for you. Marion focuses on things like the advisory function of these systems, the lawfulness, the granularity and so on. I would recommend building on those to come up with more vertical principles that are directly applicable to law enforcement.

In the final node, we need to consider where it is appropriate to draw red lines. We have talked about toxic forms of biometric surveillance. We might have also talked about other toxic forms of predictive policing. We might have talked about toxic forms of what are called emotion detection algorithms. These need to be closely considered as a precautionary principle whereby certain systems should simply not be released into the world if their human impacts will be potentially devastating.

Professor Michael Wooldridge: It is bit hard to follow that. It was a great list, a very comprehensive list, and I do not envy the people who will have to transcribe it. Genuinely, I thought that was an extremely good list, so all I will do is emphasise and underline a couple of points that David made and that we have already made.

Number one is that we need to avoid situations where people abdicate their responsibilities to a machine. That is a fundamental guiding principle that we need to hold on to. If we will use these tools in the future—we have talked about the negatives, but these advanced and AI technologies will do wonderful things as well—ideally engineer situations where it is not possible to abdicate decisions to these machines. I really hope that the committee will be able to keep hold of that.

The other big headline, and it almost sounds like a record that is stuck on repeat, is to do with data. Without data, these machine-learning technologies are not possible. They absolutely rely on data. They rely on training data. One of the defining conundrums of AI at the moment is that AI can do great things for you if you give it your data. We have not yet found our way in the world of AI, of ubiquitous data, where, as we live our digital lives, we radiate data into the world. That has been particularly the case during the pandemic, where we have done everything digitally because we could not do it face to face. Keeping control of that data and managing the way it is used to build programs that affect us is absolutely central. As I say, data sounds boring because it is not an original thing to point out, but it is absolutely central.

Those are the two points that I would highlight in addition to David's excellent list: first, particularly in law enforcement scenarios we need to avoid the idea of delegating our responsibilities—our ethical responsibilities, our moral responsibilities—to a machine; secondly, data is critical to the whole thing.

Baroness Sanderson of Welton: That was excellent. Thank you both very much.

Q37 **Baroness Hallett:** My question is probably for Professor Wooldridge. Are you confident that we have sufficiently embedded in the training of computer scientists the teaching of ethics?

The Chair: That is a wonderful question. Baroness Chakrabarti?

Q38 **Baroness Chakrabarti:** You described an algorithm as a set of detailed instructions, like a recipe, but surely a list of detailed instructions is also like a law. Given what you said about the need for transparency, accountability and democracy, is there an argument for publishing, scrutinising and authorising algorithms in certain spheres in Parliament, just as we do with conventional laws, and would that be achievable and translatable for legislators like us?

Professor Michael Wooldridge: Taking the questions in reverse order, I think there certainly is. I do not think it is realistic to take a computer program and present that to Parliament, but the idea that companies should make their code available for audit is an entirely reasonable one. To a certain extent, that happens in certain industries already. I have already mentioned the airline industry, where there are very rigorous certification processes before software is allowed to go out. It does not prevent accidents happening, as we saw 18 months ago, but it is a good system, nevertheless. So, yes, some audit process, absolutely.

On the teaching of ethics question, there has been a transformation in the last five years. I used to lecture in ethics at my previous institution, the University of Liverpool, and it is an embedded part. The students hated it: "Why are we learning this stuff? It's boring. This isn't going to get us a job". That was the argument.

That has changed completely. Students are absolutely demanding it, and let me tell you why. We use case studies, and the example I used in my lectures was, "Imagine you are being asked to write some software to illegally download MP3 files, audio files". I used the example of a Russian website. Apologies, Russia, but that was the current example at the time. That was the example of the day: downloading MP3s from a dodgy website in Russia. It turns out that a student in their bedroom in Middlesbrough with a laptop connected to the internet can change the outcome of an election on the other side of the world. That is the difference in the world we are in now, a connected world with social media and so on, and that is why students are demanding it.

We have seen a rapid change in attitudes over the last 10 years. Students now are not just more receptive but much more demanding of the ethical training. You have seen across the board, including at my own institute, a fundamental change in the way ethics is taught, not just in computing but in related disciplines like statistics and so on.

Dr David Leslie: I concur with what Mike said about the transformation in attitudes. In my last teaching gig I taught a pilot programme for teaching ethics to computer engineers. It was in the public space of the department and, lo and behold, two sessions in and the head of the computer science department had crept into the back of the room to listen in because there had been interest in it. That is the case.

On the transparency issue, I recommend the Law Society report on the use of algorithmic systems in criminal justice. It has suggested a public registry of algorithms in the criminal justice sector. On a wider frame, you should think about a public registry of algorithms at the public sector level. I think it would be important for us to consider in the sense that there should not be opacity when it comes to the use of a system. Maybe, as Mike suggested, we need to navigate how much provision of information about that system we are giving. You do not want to give the code, but you do want to give relevant information so that citizens can, in an informed way, understand when, where and how systems are being used that affect them.

Lord Blunkett: As long as we do not end up with an algorithm that evaluates the algorithms.

The Chair: Yes. As I told our staff, looking back to remind myself of a hotel in France that I stayed in a few years ago, I have realised that the restaurant is called L'algorithme, and I have no idea how they got to the menu that they proposed.

I want to thank our witnesses very much. If there is more that you would like us to think about following today, we would be delighted to have the thoughts. I have takeaway questions for you, following the last exchanges. Who should be the regulator? How would you design the terms of reference for a regulator? Thank you all very much indeed.