# Committee on Standards in Public Life Roundtable: Review into Artificial Intelligence and Public Standards 23 May 2019

Committee on Standards in Public Life

*Lord Jonathan Evans, Chair*

*Dame Shirley Pearce, Member*

*Jane Ramsey, Member*

*Monisha Shah, Member*

*Amy Austin, Policy Advisor*

*Aaron Simons, Policy Advisor*

*Lesley Bainsfair, Head of Secretariat*

Attendees

*Oliver Buckley, Executive Director, Centre for Data, Ethics and Innovation*

*Fiona Butcher, Defence, Science and Technology Lab, Ministry of Defence*

*Bethan Charnley, Innovation Policy Lead, Government Digital Service*

*Jimmy Elliot, General Counsel and Head of Legal, SAS Institute*

*Sabine Gerdon, Project Lead, Office for AI, DCMS, and World Economic Forum Fellow*

*Sana Khareghani, Deputy Director, Office for AI, DCMS & BEIS*

*Alex Lawrence-Archer, Head of Strategy and Governance, CDEI*

*Simon McDougall,* *Executive Director for Tech, Policy and Innovation, Information Commissioner's Office*

*Ian O’Gara, Digital Strategy Director (Public Sector), Accenture*

*Marion Oswald, Senior Fellow, Department of Law, University of Winchester*

Lord Evans

Good morning. I would like to welcome everybody to this, which is our first roundtable in the inquiry we are doing into ethics and standards and AI. I am Jonathan Evans. I am the Chairman of the Committee on Standards in Public Life and I have with me a number of colleagues from the Committee: Jane Ramsey, Shirley Pearce and Monisha Shah. We are very pleased that you are here. We also have a number of colleagues who have been organising this and I know with whom you have had a number of conversations and they have done some preparatory work. I think everybody has received the bundle of papers from which we are going to be working this morning. We are very conscious that we are under some time pressure, because we only have two and a half hours or so and quite a big agenda. I hope, nevertheless, that we will be able to cover that ground effectively.

As a very brief introduction for those who are not familiar with the Committee on Standards in Public Life, it is a very strange creature. We are independent of government, we are not a regulator and we are not a tribunal, but we do advise the Prime Minister on ethical standards in public service, which can include those elements of the private sector that provide public services. We work principally by looking for areas of actual or potential concern in the standards landscape.

At the moment the one we are in the relatively early stages of looking at is the question of AI, because we recognise that the public sector is likely to be increasingly supported or driven by AI capabilities over the next 20 years. As the system is starting to come to terms with that, we thought it would be useful to look at the question of the extent to which the standards that public service is committed to, particularly the seven principles of public life, are going to be challenged by AI and the extent to which the structures and arrangements for the use of AI in the public service are taking those standards into consideration, with the hope that in 20 years time, when this is much more mature, we will still be able to have confidence that organisations and individuals operating in the public service are able to meet the standards in the way that we would all hope. In order to do that, clearly that needs to be deliberated and there needs to be a decision on it, so we are investigating and looking at the emerging use of AI in the public sector and also the structures and arrangements in place to underpin the ethical standards that need to be in place.

To do this, we have devised and articulated six standards challenges built around some examination of the seven principles of public life, and it is those that we would like to discuss in the course of the roundtable today. Those challenges are not set in stone. It may be that, as we discuss it, we say, ‘Actually, that is not such a big one’, or, ‘We have missed something here’, and I think that will be a very valuable takeaway from our point of view.

I am not going to invite everybody to give a formal introduction, because I think by the time we have done it we will probably have used up about 20% of the time, and I think everyone has seen in the papers who is present. We have a really good selection of people from across government and involved in supporting government, so that is valuable. We are having another roundtable shortly with academics and the various roundtables and research we are doing will help us to get our heads around the questions that we need to address.

The roundtable is being transcribed and a copy of the transcript will be sent to participants for checking after the event. In due course, that will be published on our website as part of our transparency. The roundtable is also being live-tweeted using #AIPublicLife, but individual comments will not be attributed. Just for those who follow, it will be of interest. That is probably all I need to say by way of introduction. Thank you very much for taking part; I look forward to the conversation.

We have broken the whole roundtable down into two sessions. Shirley has very kindly agreed to sum up at the end of each session, so that we can take a breath and work out where we think we have got to. At the end we are going to have a wrap-up, which again Shirley has agreed to lead, to just review whether we think that the issues have covered the challenges that we need to consider and whether there are any issues that have come out that we need to take further. That is the format.

You have been allocated biscuits on the basis of extensive algorithmic work and we have worked out which ones you like most, which is why some people have chocolate, like me, and some people have lemon, like Shirley.

We are going to start with session 1, which is divided into various parts. We do not have to cover everything. If we get to a point where we have had a really useful, productive and illuminating conversation but have not answered question 6(b) or something, that will not be the end of the world. It is more important that we get to issues that really matter than it is that we fulfil the various nooks and crannies of the agenda.

I wanted to start with this question of responsibility and, in a sense, accountability. We saw one of the standards challenges being the balance of responsibility between the humans involved in the process and AI as a technology helping or making decisions in the public sector and the need for clarification and definition around that. This is important in our view, because we need to ensure that accountability is not undermined when AI systems are brought into government and we do not want a situation where public officials can say, ‘The computer made me do it’, and there is not any way to get behind that, because I think that would be a very bad outcome. There are a number of questions around that, which I think you have had in the pre-read pack.

We might just start with the first one there, which is whether there are any thoughts that people have or comments you want to make on this question of how responsibility for AI systems can be clarified. Are the human-in-the-loop models useful? What does that mean? Can we define any individual systems? Is it viable to say of a particular system, ‘It only has these limits’, and that, ‘This cannot be used for definitive decisions’? We would very much welcome comments and contributions around that area.

Fiona Butcher

From an MoD perspective, we have been looking at this in regard to the concepts and doctrine around the use of this technology in supporting future decision-making and one of the concepts that we have defined is ‘human-machine teaming’, so rather than thinking about this as ‘human in the loop’, thinking about it as a partnership between an intelligent machine and a person, and really trying to understand how you can get the best from both. With our systems today, we take a very human-centred design approach. This is about supporting the human decision-making, rather than saying, ‘Is the human in the loop, on the loop or off the loop?’

Lord Evans

In that partnership model that you talk about, would you see different levels of human judgment being exercised for different issues? What sort of thought have you given to how that would operate in practice?

Fiona Butcher

When thinking about the human control element, it is about thinking of the whole system from design, to procurement, to use of an AI system. I know a couple of your questions refer to the procurement of this technology and the human decision-making that is involved at that stage. We would suggest that there is an element of human control at every stage including defining specifications, so referring to your question about which elements it is appropriate to automate using this technology and which elements would still need to remain within the control of the human, it is very context-dependent.

Lord Evans

Just to clarify on that, conceptually that sounds very sensible. Have you gone beyond, as it were, that level of thought and started to try to codify or to provide guidance as to how these questions are answered, or is that something that you will be moving to in due course?

Fiona Butcher

These are the sorts of questions that we are exploring within our research at the moment.

Lord Evans

That is interesting.

Simon McDougall

One of the things to keep in mind, if and when we become more dependent on AI systems, is the split between individual responsibility for the human that is involved in that process and overall accountability for the organisation for the AI that is being used. It is an issue we have now, but it becomes exacerbated with AI because there are more design decisions and decisions around what data to acquire. There are bigger decisions being made prior to the operation of the system that affect how the system works and how the decisions work. You need to say that a lot of the accountability shifts onto the organisation and shifts onto senior management, however that is constituted, and sometimes that means it goes away from the humans who are involved somehow. That is not to say that the humans have no responsibility at all. That depends, as Fiona says, on the context, but there is a balance there.

The other thing that shifts the balance in that direction is, especially once we have effective AI supporting decision-making, there is plenty of research out there that highlights the very human trait to get a bit lazy quite quickly. If you have validated an obviously correct decision 19 times out of 20 and the 20th one pops up, you are very unlikely to start to move against it, so we are going to have to modify our expectations in some contexts for what responsibility the individual could be expected to take.

Lord Evans

In saying that, does that mean therefore, that we need to underpin the idea that people must take responsibility or that we have to accept that they will not be in a position to take responsibility?

Simon McDougall

I personally think very often it is going to be the latter. Thinking of the obvious one, self-driving cars, even when people’s own lives are at stake they tend to start switching off a little bit if they are whizzing along and the car has been going fine for the last few journeys.

Sana Khareghani

I may become a broken record by the end of this, but it really does depend on context when you talk about this. I totally take Simon’s point on losing some of the human aspect of error finding if you keep getting something with an enormously low error rate, but there are ways around that, and context-specific ways around that. If you think about airport security, for example, a lot of stuff that goes through the scanners is picked up by the computer and machine vision algorithms, but for various reasons they will insert problems into that to ensure that the people who are working with these systems stay vigilant and continue to be trained.

I totally agree it is about teaming. It is not about inserting somebody in the loop, around the loop, on the loop, behind the loop or anything like that. It is about redesigning that end-to-end process to say, ‘How do these two systems, whether it is the brain system, the human or an algorithm system, work together to get us much better results?’. We have seen, again context-specifically, humans by themselves are not so good and a machine by itself is not so good, but put them together and you get a superhuman. That onus is on these use-specific cases to figure out exactly what that end-to-end process looks like to give us that superhuman capability.

Simon McDougall

We have had that concept in the privacy world for a long time now around privacy by design, and there are lots of practical mechanics as to how that kind of thing works. There is plenty of good thinking about how to integrate that into this new world, which is obviously way beyond privacy, but which has some similar risks, and imbed that into the design process.

Fiona Butcher

By taking a whole-systems approach you can try to mitigate the risk of an accountability gap, where you somehow assume that the person is no longer accountable. If you are saying that there may be some decisions that need to made so rapidly that the machine makes the decision (if it has been appropriately codified), there is still human accountability at the design stage and in the verification and validation of the AI system before it is put into use. This means you do not have an accountability gap as ultimately a human is still accountable.

Sabine Gerdon

Especially in the design process, it is also important to bring in multi-disciplinary teams to perform on all sorts of backgrounds and to guarantee diversity, but also diversity of thought and diversity of accountability to design the process correctly and ensure that the responsibility is somewhat or shared or those decisions are made and signed off with the people who have something to say in the deployment as well as in the design.

Fiona Butcher

There can also be a question around the competence of the decision-maker. One of your questions is about whether responsibility would be deferred to senior managers within an organisation. They would need to have the appropriate advice for those sorts of decisions, because some of this technology is quite sophisticated, with the greatest respect to senior managers in organisations. They would need that appropriate advice before making those decisions.

Simon McDougall

That is a very practical challenge at all levels, because we need to be realistic about this. Sana, your example of the security machines at an airport is great, but they are very highly trained individuals and there is a very clear risk they are managing. What we are going to see as AI becomes more ubiquitous is less motivated and less trained individuals being these humans in the loop, and I think a lot of the time the errors are going to come from underestimating the need for training if you are going to place some reliance on humans plus AI.

Lord Evans

Instinctively I agree with all this. My slight anxiety is that it has a slightly elite feel to it, in the sense that at an individual level the person who is the recipient of the decision will not be able to explain, but if they happen to be the sort of person who is interested in the governance of organisations then they will be able to get a really good explanation. Most people are not and that will not feel like accountability if they say, ‘Look at the way in which we govern these risks’.

Simon McDougall

We can share our results with the Committee. We are releasing our first interim paper next week, but we are doing work at the ICO with the Alan Turing Institute on explaining AI decisions. Let us not live-tweet this bit, because it is meant to be the key point of the meeting next week. Without giving away all the findings of the report, we are running citizens’ juries with real people out there, and one of the things we have found is in terms of the explanations that they want around different AI-driven decisions, it is very context-specific and the level of granularity they want and the level of human intervention they want varies massively between, say, a medical decision around their health versus, let us say, a decision around their benefits. On a very rough sliding scale, the more faith they have in an institution, the less explanation they actually want. If it is medics doing it, then they trust the medics and they need less explanation.

Lord Evans

Are there other people who want to come in?

Marion Oswald

There is an interesting point about what people actually want, and we need to consider what the decision-maker needs in terms of information to make sure that that decision is a legitimate one, and that, if they have to exercise an element of discretion, the tool is not somehow fettering that discretion and meaning that they are not looking at other factors that they should look at.

I have had an opportunity to have a look at the HART tool that Durham police have rolled out, which is very clearly badged as a decision support tool. The custody officers are told that they can look at the output of this tool, but they also must consider all the other factors they need to consider in making a judgment about an individual. If anyone is interested, I have a copy of a screenshot of the way that that output is shown to the custody officer, and we may or may not think that it is giving the information that that custody officer needs. I think one of the issues is the ‘traffic-lighting’ of the output. The one I have is green (for low risk), and the officer is told that ‘You must go and look at all the other information’. What is that actually telling the officer about that individual?

The design of these tools is so crucial to making sure the decision is legitimate and that is something that has come up in the West Midlands Police Committee that I am chair of. The real key, as others were saying, is the context, how that decision is going to be used and how it is going to fit in with the decision process. I have been to the States recently and I had an opportunity to speak to some people within the legal aid arena and the District Attorney in New York. The culture over there has resulted in police officers not challenging these decisions and not going against them, and I think that is a big risk, depending on the way the processes and procedures are produced internally.

Lord Evans

Do others want to come in on these points?

Ian O’Gara

For me, it starts with that question of what AI is, because it means many different things for many different people. For some people, it is just a bit of RPA process enhancement object recognition. For other people, it is right down the end of predicting who is going to be doing it without any influence of humanity or data. We need to probably start to define what we mean and look at the gradation of those definitions.

The project I am working on, which is looking at utilising data and creating centralised analytics capabilities for law enforcement, goes from the point of data engineering, data matching and network analysis for working through OCGs (as an example), through to prediction of people who have a propensity towards a certain type of criminality or violence. This is just within the proof of concept at this point in time. Before we would even consider deploying it, we would need to work through all the implications, as Marion said, about how it will be used.

The context matters, not only in how it will be used and what the interventions are that people will be doing with it, but also how it will fit into the process and what the use case is that you are trying to achieve. You talk about systems thinking. We need to look at this in the round. As a supplier in these types of worlds, for me and for my team, what we need or what we would request from organisations such as this and from government is a set of standards. I have worked with the GDS data science framework a couple of years back and I quite like that, because at least it starts codifying standards. We need to look at how we can start to bring these together. You talked about how to ensure that people stop just relying on the output, and I think we are going to come on to the training and the culture. This is not an easy issue and it is an interesting process, but for me it is the definition, it is the use cases and it is the contextual end-to-end piece.

Lord Evans

As a supplier to the public sector on this, is it your experience that a client is well-acquainted with these issues in so far as the ethics are concerned and able to give you the guidance you want, or are you coming to them and explaining issues that they were not quite so familiar with? We are not talking about any one client.

Ian O’Gara

I would say clients are starting to get more and more attuned to these situations and more and more bodies are becoming more interested in it. We are learning together; that is probably the best way to describe it. It is definitely not part of the procurement process at this point in time, but I think as people start to work through the issues, it will hopefully become part of that.

Simon McDougall

I would disagree with you, Ian, on one point there, for the purposes of this Committee. I agree with you on the standards and your point that overall we need to be very careful how we define AI. In a way, for this Committee I am not sure whether we really need to delve into the definitions of AI, because almost the challenge we are facing is that, in the medium term, the amount the data we are going to collect is going to continue to increase rapidly and the ability of technology to analyse that data is going to increase rapidly, even if actually the underlying techniques around machine learning are not really changing very much, but just the computer power and amount of data means that we are getting new uses of it. We are going to be in a world where there are going to be far more techniques available to help us make decisions about things, and that to me is the underlying challenge we have here. AI is just a mechanism for that right now. The challenge we have in terms of the principles in terms of this Committee is how we have to consider the challenge of us being more reliant on machines to make these decisions.

Lord Evans

So far we have tried not to get into the technical definition of AI, partly because I probably could not understand it but also because there appear to be different definitions. Even in our conversations with the Alan Turing Institute, they sort of said, ‘This stuff’, rather than ‘AI’, so they were clearly wanting to maintain a bit of leeway as to where they felt the challenges emerged from. Jimmy, do you want to come in?

Jimmy Elliot

Yes, I had a couple of notes. There is an interesting conversation around the controls that you can put in place to help give some confidence around responsibility and decision-making. You have probably read it, but the EU guidelines on AI and ethics have a really useful approach where they are piloting an assessment list matched to what their principles are. That kind of approach can help, because it is different by context, like you said. That is quite neat to look at. Before you can get into designing the controls, the first question for responsibility seems to be, ‘When are we comfortable with the balance that we have between selflessness and public duty versus human autonomy and dignity? Where is the balance? When do people need some guidance, perhaps on the areas that we have decided are red lines and we are not quite so comfortable going there at all without there being perhaps more of a view?’

Lord Evans

Oliver, did you want to come in?

Oliver Buckley

Yes. Essentially, we are all in pretty violent agreement on some of the fundamentals here. I think that it is interesting then to reflect on the language when we talk about the balance of responsibility and when we talk about humans in the loop. I think what we are all saying is that this is not actually about a balance of responsibility between humans and machines. Ultimately we are very clear: humans are responsible.

There is a question about among the humans who are responsible, what degree of responsibility do they have, whether you are the person at the frontline responding to the advice you have had about an individual case or you are the senior accountable officer for the system? Those levels of responsibility may vary, but humans are responsible; that is the first point.

The second related point is just I think that this human-machine-teaming language is important, particularly in terms of what the fact that we have come to talk about ‘human in the loop’ reflects. I think it reflects sometimes a sense that the AI is the thing we are looking to bring into the system and we somehow have to fit humans around it. I think that sometimes reflects a culture around technology in government and public services, where there is a sense that innovation is key, being open to the new and novel is key, and what you need to demonstrate is that you are embracing new technology first and then figure out how to make that work second. It feels like we have that the wrong way around. The starting point always needs to be what the system is that we want to design and the outcomes that we want to see and then the technology has to bend and fit to that, not the other way around.

Sabine Gerdon

Just to follow up directly on Ollie, even before that we need to ask the question, ‘What is the problem we are trying to solve and is AI really the right solution to solve that problem?’ That is one of the key starting points that I start with when I am looking at procurement guidelines for AI. Part of the project that I am leading on with the World Economic Forum and the Office for AI is to look at how we can, within the procurement rules that we have, ensure that we, more effectively but also more responsibly, adopt AI as the public service.

What is interesting and what has not come up in the discussion is also the responsibility on the supplier of the technology, because in our current framework this relationship stops at the moment when we set the contract and in the current procurement process it is not thought through how we can have this much closer relationship with the supplier that we need to ensure that the staff that interact with the model are actually well trained and continue to be well trained, as well as the fact that the model needs to be maintained over a long period of time, because it is evolving constantly. This relationship needs to be much closer than it used to be, so, therefore, in redesigning a new procurement system or process or thinking of how we integrate that in our current system, I think it is really important to think about exactly those points that Oli mentioned before. How can we support officials in making those decisions by having a well-designed process where they are reminded at every stage of what consequences the decisions might have and what questions they need to ask suppliers to make more effective decisions?

Lord Evans

We could carry on on this because it is jolly interesting, but we had better move on to the question of data bias, which I think is one of the most publically discussed issues in this space. How big a problem do we think data bias is in AI-enabled or AI-assisted public services? How confident are we that current efforts to combat data bias are fit for purpose?

Simon McDougall

It is a big problem and we are not confident.

Lord Evans

Right. Moving on then… [*Laughter*] It is interesting, though, because it is much discussed. One would imagine that people had probably given more thought to this than some of the other areas that are less obvious problems in this space.

Simon McDougall

I will put a marker down, because otherwise we might forget it. Before we dive into the issues around bias in decisions, in the future we will have an AI process replaced with another AI process, but we have to keep in mind that right now we are more likely to be replacing a human process with an AI process. All us humans are bringing a whole suitcase of preconceptions, prejudices and baggage along with us to that decision, some conscious and some unconscious. As we talk around bias in AI – and there is plenty of stuff to talk about – we have to keep in mind we are not moving from a beautiful neutral model. We need to balance these things up. Very often, even if your model is not perfect, it may be less imperfect than what you are replacing.

Lord Evans

Ian, is this something that as a supplier you give a lot of thought to?

Ian O’Gara

A huge amount. We have tools, [thought programmes?] and work going on to try to work out automatically whether the data we are ingesting, not only within the public sector but elsewhere –mortgage applications, insurance and banking – is perpetuating any bias that is present within the system or within the datasets that we have. We are always trying to strip out any level of bias that could be present in the data.

Within the proof of concept that we are working on at the moment, we are stripping out any questions around ethnicity, race or geography, so we are not using that as a proxy of location of individuals to try to test through what that means. Will bias still exist in that data? Yes. I would question whether we would ever be able to strip out all bias, but you will certainly be able to look at the balance of data that you have against certain demographics to try to work through whether one group is over-represented or less represented.

I am going to be a bit of a pain when I talk about the definition of AI, because for me it is quite a challenging question. When you talked, Simon, about the replacement of a human process with an AI process, that is definitely part of the process, but sometimes we are replacing no process with an AI process. I was with a union yesterday to talk through a workforce wellbeing model that we were creating to look at instances of low wellbeing within forces and they had done a similar piece of work to the one we had done, but they had looked at a very small dataset in one location to try to work out what was going on and why there was a higher incidence of sick leave. It took them 10 months in a very paper-based exercise, whereas we were able to do it much more quickly than that.

The data is biased. We cannot get beyond that, so we are going to have to find ways of recognising, flagging that, addressing that where possible and creating learning processes back into the system, so we can start to address those issues. I sometimes see others layering on other black box impenetrable scanning systems that give you a view of the biases in the data but do not really help that clarity come through. We should address it, but we should look to see how we can address it effectively.

Sana Khareghani

My fear or reticence changes on a daily basis with this topic. I should just preface that by saying that I am probably the most enthusiastic and optimistic person about this. I often get told by people like Oli that I need to calm down. Having said that, I do go through waves of what worries me the most, and at the moment the thing that worries me the most is this concept of taking the bias out of datasets. That really worries me. I am not saying that datasets are not biased, but I am asking how do know that you are not making a dataset less biased? What are you taking out? I understand criteria around protected characteristics and making sure the algorithm is not using protected characteristics, but I do not think human cleansing of data is the answer. That is an ultimate worry for me, because then whose choice is that? Is it your choice? Your company is making the decision that this is now less biased. What about my decision about whether or not that is true? What about you bringing your own biases, as Simon pointed out, into that?

I do agree. I think this is a multi-faceted question. Ultimately, there is a way of getting things better than they are. This may not be the right option, but there are ways of joining datasets and ensuring that there is a more diverse dataset by just having datasets that come from multiple sources, making sure there is a diversity of thought in the creation of the algorithms, making sure there is a diverse set of data and you are using data-sharing models.

These will get us closer to taking out some of the bias, but also, as we have talked about a lot, the design is enormously important. We are aware, for example, that there is bias in our current systems in hiring people and these are examples that are out today. If you look at your hiring data on who has succeeded in your company and then use that to find new candidates you are going to be biased in the way that you hire employees, but that should show us that the decision-making criteria we have been using has been wrong and what it should help us do is design algorithms that merit multiple facets of somebody’s characteristics.

Lord Evans

Bethan, you wanted to come in.

Bethan Charnley

From a GDS perspective, there are probably two things that have come up in our work. First of all, a year or so ago DCMS created the data ethics framework, which is a really great tool for helping service teams think through these questions. The consideration we have is that at no point is that enforced through the process of assuring public service projects. It is trying to think about how you might go about integrating into the GDS spend controls or standards assurance processes and whether, indeed, that is an appropriate thing to do. Do we want to say that ethics is something we can rubber-stamp on projects and say, ‘Okay. This has been done,’ or is it an ongoing process that needs to continue and it is more a case of saying, ‘Have you considered these things as part of that process?’

The second thing is when we talk to data scientists about this at GDS, it is clear that, although in the context of AI there are new challenges, data bias as a concept is not new. Statisticians and analysts have been dealing with this for a really long time and within government we are quite quick to see data scientists as a separate profession over here and ignore that we have a huge statistical and analytical capability, with years and years of experience and training in dealing with these kinds of things. There are huge resources. I think there is an Aqua Book, which is government guidance on how we should remove data bias outside of an AI context. I think it is also a case of bringing those existing resources and that existing knowledge into the conversation as well.

Fiona Butcher

I think those are really important points actually. In some senses, this is not new and we just need to apply some of the good practice that we already have around experimental design and incorporate this into systems design. I would also echo Simon’s point about the importance of diversity in those people that are making some of the decisions around the design and implementation of AI.

I have just one other final point with regard to the language of substituting people processes with AI processes. Again, we try not to think about that ‘substitution myth’: that you can take out a human and just put out in a computer. You are, again, required to think of this as a whole system approach and focus on the problem that you are trying to solve.

Lord Evans

Marion, you wanted to come in.

Marion Oswald

As a lawyer, I sometimes think we are talking slightly at cross-purposes when we talk about bias as well, in that when I think about bias I am thinking about potentially somebody with an underlying motive, conflict of interest or an outcome that results in illegal discrimination. When bias is referred to in other contexts, it sometimes means statistical bias. Effectively, all data is going to be in some way biased, because it is a subset of the information that is relevant to that particular scenario. We could do more to help understanding of those different meanings of bias and how they interrelate with each other, because you could have bias in the inputs, in that they are somehow reflective of a circumstance that was inappropriate at the time. You could have bias in the outputs, in that they do not reflect the scenario that you are looking at – for example, tools trained on training data relating to male offenders and you are actually dealing with a female offender or a young offender, so that is totally the wrong tool to be using – or the whole output could end up with some sort of discriminatory result because of the way it is being used.

As I think was said earlier, it is not just as a case of stripping out certain factors. It is also looking at what outcomes are going to happen and whether those could be discriminatory or not. Some of those factors are going to be very relevant. A tool we have looked at at West Midlands Police had all intelligence information built in as input factors, including things like number of stop and search counts, and that raised red flags around what that could be a proxy for in that particular region. It is very important I think to look at the relevance of all the input data. Are they relevant to the outcome you are trying to achieve? Do they have a fair connection?

Simon McDougall

I have a quick observation. This is going back to the data bias, so apologies to Marion. One thing I think we need to keep in mind is we often focus on bias being something when we are setting up these new systems on the training data and the initial design, but over time our values may change and maybe not the principles themselves, but underlying that what we think is acceptable and unacceptable in terms of types of discrimination and the way we identify people and understand people may change. This has to be a living process. We cannot just conclude that we have built an AI system and it is not biased at the start, whatever that means, and then away we go and it is running for the next 20 years, because I think we would be quite appalled at what we saw by the end of it.

Oliver Buckley

That is why I think this comes back to the first discussion, which was about accountability and how you ensure appropriate transparency sufficient to ensure that accountability. I am just looking again at the language around the challenge. There may be situations where you want AI to reflect real world demographic bias, in order to understand it. I think that perhaps at the level of a principle what we might want to say is it is unacceptable not to know the ways in which your system is biased, and you are then required to account for how you use and understand the results of that system in that context. You need to be able to provide a justification and that justification has to be subject to scrutiny and challenge.

Lord Evans

I think that so far, from our side, the way we have been using bias is probably only part of the meaning of it. We are thinking of it in terms of, to use the pejorative term, relating to prejudice, rather than the fact that this may not be all the data that is available. I think we need to be careful about how we are using the term. Okay. That is good.

Let us move on to the question of ensuring that AI is used in the public interest. We are conscious that AI is powerful, which is a good thing if used correctly and a bad thing if not used correctly, and it brings risks therefore. Ensuring that those risks are recognised and it is used in a way that is in the public interest is an important aspect of this. One of the underlying themes from quite a lot of what our Committee looks at is the abuse of power in a variety of ways, for personal enrichment, for bullying and harassment or whatever. This is another of those areas where power is being entrusted to institutions or individuals, which is going to be a slightly different sort of power potentially over the coming years. How do we ensure that that factor is fully taken into consideration as we procure and deploy new technologies of this sort?

So far in the initial conversations we have had with a few companies, I do not think we have yet come across anybody who says, ‘Yes, when we are talking to those who are procuring our services or products, then standards are definitely in that conversation’. So far everybody has said, ‘No, I have never had a conversation with anybody that included standards. It was all about effectiveness.’ ‘Is it going to work? Is it going to save me money? Yes, I want it.’ It is not surprising, given the pressures on public sector budgets, etc, but it did strike me as being, at the least, a lost opportunity and potentially quite worrying.

Sana Khareghani

I would add that there might be another reason. We are still not at a place where this is general purpose technology that you can buy off the shelf, in which case there would be standards around it. Please come in if I am wrong.

Lord Evans

Just to clarify what I was saying, when I use the word ‘standards’ in the context I guess I am meaning it in terms of the standards in public life, so that is my personal bias, as it were, rather than performance standards or more traditional industrial standards. If there was ambiguity on that I apologise.

Sana Khareghani

Then, yes, I think you are right. From that perspective, are we putting the citizen first, in terms of when we are employing this, or are we putting the user first? Bethan can come in from a GDS perspective. In our procurement guidelines the user need is the number one thing we put in, whether that user need is the public or the civil service themselves that are using this thing that we are procuring.

The case where this becomes a bit fuzzy when it comes to AI solutions, again, depends on what we are talking about. Are we talking about an automation? Are we automating a bit of services, in which case we kind of abstract away from the end-to-end process and do not look at it from a systems view? We think about that specific thing, so we stop talking about that standard as you would define it.

If we are thinking about the full thing, where we have an issue and we want to use AI to solve it, it becomes such a bespoke solution to a challenge that ultimately people do not specifically say they are thinking about the standard, because obviously the challenge is improving standards if that makes sense. I wonder whether actually we are coming at this the wrong way, in terms of saying this is not being thought about, but that is I guess where I am coming from on this.

Ian O’Gara

I heartedly agree. There are not standards. When I talk about standards I talk somewhere around procurement standards, somewhere around the actual technical standards and somewhere around the actual ethical principled standards. When we step into these arenas we are trying to navigate all of those elements, not as part of the procurement process, but as part of the engagement process. It is the definition of the problem, the design of the solution and the thinking through the interventions. It is a very iterative process to try to work out who all the stakeholders are who we need to consider and then how we include all their thoughts, their views and the appropriate governance structure around what it is we are trying to do.

I will come back to that definition of AI again and again and again, because are we looking at doing process automation or are we looking at prediction of criminality within a certain network, which have very different dimensions, very different tools and very different data? All these things need to be considered and it is important for us to consider them, because it is very hard to codify them at this point in time when there are lots of moving parts around the pieces. It is something we want to do and, thus, with the West Midlands Police we are looking at instituting ethical standards and ethical oversight of the work that we are working to ensure that has been taken into account and to work through what the practical process steps are that we need to think through as we are going through the design process with data scientists and going through the consultation process.

Oliver Buckley

I guess perhaps one area in which AI perhaps is presenting new kinds of challenge in this situation is that by their nature these systems are dynamic and you might not fully understand the power of the system at the point that you procure it. That power may change, so you might not know how effective it may prove to be and then, therefore, how proportionate it ultimately is for the task that you have set it. I wonder if, therefore, what we need to move towards is a more iterative view of the decisions and governance around this stuff. Perhaps with other types of systems, where you know exactly what it does is what it says on the tin, that can be a one-time decision subject to periodic review. Perhaps we need a more dynamic system of review in recognition of the dynamic ways in which these systems operate.

Lord Evans

How does this look for SAS? Do you know?

Jimmy Elliot

We are in the middle of debating this, a lot like a number of tech companies debating where we should be helping customers and where the areas are where we are uncomfortable where we think it is not appropriate. With other companies, shareholders seem to be making that decision. With others, it is employees. I do not think we have figured out the best way to do that right now, but my personal feeling is it should be less for the supplier to decide whether they do not go, because the worst thing then for the government is that all the way people you can work with is narrowed to those that perhaps have not thought through the ethics of it quite so well. I would like that to be more of a partnership around in the high risk areas I would like to see SAS and other vendors talking about, ‘If we are going to go into certain applications and use cases, what does responsible selling look like for us?’

There must be more than just bringing technology. There must be something around that rounded package of the people who will interact with you to give you that diversity of view and the ongoing commitment, so we are not going to deploy and scoot off. We will to some extent have then an ongoing responsibility to take on more risk as part of that if we are going to be responsible vendors. That is the debate that is going on on the tech side. I think it is more for the governments then to try to agree common areas where they think it should not be used at all and then the partnership comes into the areas that we feel comfortable.

Fiona Butcher

Just coming back to this particular standard of AI in the public interest, we are very mindful of that in the context of defence. We have started to adopt the government’s Data Ethics Framework and we are doing some of our own internal research to understand whether there is anything within the defence context that needs to be considered in addition, because obviously we are also recruiting large numbers of the British public to be part of MoD as well. Reflecting the ethical and moral approach that the public would expect is important as an employer as well.

Lord Evans

I think it is probably fair to say the intelligence community has been having the same conversations, but they will not be taking part in these roundtables.

Sana Khareghani

I do know that people are considering ethics, but what is missing right now is that it is not codified. It is left a little bit to that conversation that is had by the people who are procuring and the responsibility is left with the tech companies for doing the development. It would be good for the Centre for Data Ethics and others to help us set where this should be codified in the conversation, so that we can either tick as a box and say, ‘We have done that. We have had that conversation or it has already been imbedded.’

Lord Evans

Is there a process heading in that direction?

Simon McDougall

As I was going to say, there are plenty of codes out there. If anything, there are too many codes out there.

Sana Khareghani

That is the problem.

Simon McDougall

Absolutely, yes.

Marion Oswald

They are not grounded in context. A lot of these codes are very high level, vague principles that do not give any practical guidance. I am very biased, but the ones that are working better are ones that are grounded within a public sector body, so you can reflect the ethical code to the function of that body, look at the statutory purposes of that body, etc, and relate what you are doing to what the public body should be doing in the public interest. Therefore, it is easier to then think about the way the AI tool is fitting into that.

Lord Evans

I will ask Sabine to come in and then Bethan. Then I am going to suggest that Shirley might like to do our little pit stop and we will go from there.

Sabine Gerdon

I wanted to emphasise that it is also not currently embedded in the procurement process. Those discussions are made, but it is not mandatory in the procurement process, nor is it mandatory in the evaluation of the procurement process. What we could do is make it a quality requirement, similar to quality requirements we already have in the specification, to have those ethics considerations as quality criteria, to also then ensure that the ethical suppliers or suppliers that think about ethics and build that in their system have a competitive advantage compared to other suppliers. You could then reflect that, depending on your procurement context, and then use that as evaluation criteria when deciding which supplier should be deploying the solution.

Lord Evans

Instinctively, I think that is a really important point that we need to align the incentives, so it is actually advantageous both to the procurer and the supplier to be seen to be meeting high ethical standards and to have the wherewithal to have that fluid and continually engaged conversation. I think if we can do that, then apart from anything else one potentially could have a premium because you are really good at doing this, as opposed to some kind of backstreet firm that is not very good at it and you can get it at a knockdown price because it is unethical. That is my caricature.

Bethan Charnley

Just coming back to one of the points that Sana made earlier about the importance of framing it with user need and starting with a problem, GDS and the Office for AI have been working on creating guidance for using AI in the public sector, which will be published shortly. We have been very careful to take that approach throughout the entire guidance that this is always about starting with user need, always about starting with the real problem and then assessing if AI is, indeed, the right technology to solve that problem with. I think we need to be careful about balancing that with the other side of things, which is bringing AI into government is a huge strategic opportunity that we need to make the most of, giving leaders the confidence to know when it is appropriate to use AI and when it is okay to say, ‘No, actually, this is not an AI project and we are going to use a different technology or a different approach for solving that’, making sure that, equally, the incentives within departments are not aligned towards building huge AI capability and doing the most AI projects if they are not necessarily the right projects to be doing and just making sure that we keep balancing that conversation.

Lord Evans

Go on then. Say what you want to say. I do not want you to burst.

Ian O’Gara

There are perverse incentives to do that. I have seen it with a number of clients where they say, ‘We need to look at our high risk individuals and screen them for warrants outstanding. Let us use AI for that’. Actually, just clean up your data a little bit and try to match nominals, and you are going to be fine, because you are only dealing with 100 people and not 10 million. Thank you.

Lord Evans

Thank you very much. Shirley now has the unenviable task of summing up the discussion.

Dame Shirley Pearce

Aaron has been making really useful notes here and I know Amy has been thinking about it too, so I am just going to do my personal impressions and bits that I have missed you can add in please. We started off with this concept of the human in the loop, on the loop, off the loop and I think there is a strong message that that is not an appropriate concept really. It is a partnership. It is a human partnership. We went on to say in that partnership it is the human that takes responsibility, so we have to be clear about that. It is not a partnership of responsibility. It is a partnership of development and delivery, but in terms of responsibility, which is what this issue is about, it is the human.

The responsibility relates to a number of different factors and I think there is a link across the three things that we have talked about that there is responsibility for a raft of different factors at lots of different stages. Ethics and all these other things need to be at each of the stages of design, procurement and operational delivery, so you cannot separate it out. There is a very strong message about depending on context. We are talking about AI as if it is all one thing. We have not got a definition. Maybe we do or do not need a definition, but we certainly need to recognise that it is not one thing. It is entirely dependent on context.

We then moved on to bias and I think we all agreed that, whether you are a machine or a human, bias is there. It is a matter of being maximally confident that you have identified those sources of bias and that the impact of the bias will change over time, because these are dynamic systems. We need to have a confidence that in the design, the delivery and the procurement you have as diverse a range of views, people and perceptions as possible.

I also got the message that some of these bias issues are not so different for AI than they are in the huge amount of data stuff that goes on anyway and that there are guidelines around. Perhaps we ought to be more sharp at recognising those and ensuring that they are fit for purpose for the kinds of AI decisions, so I thought that was rather helpful.

I think I heard someone say there is a data ethics framework, but it is not being enforced and we need to think about how we can make it enforced in a more useful way. Somebody said it is unacceptable not to know how it is biased. Was that you, Oliver? That made me think, ‘The trouble is with a lot of the unconscious bias you do not know that you do not know’, so it is not quite as simple. Having made every effort to ensure that you understand potential sources of bias and what you might [be doing?] is as essential for a human delivered system as it is for an AI delivered system.

I think I picked up that there is quite a bit of work on the public interest ethical issues for perhaps us to be doing, in terms of articulating what should be an ongoing partnership between providers, procurers and delivers of the service, which are constantly assessing and thinking about their role in the delivery of an ethical service to the public. It would be good if we had greater clarity about where and how ethical issues are discussed at each stage of that process. I think I heard that too many codes are not grounded in context. How we help the system as a whole ground ethical codes in practice and link them to incentives is a challenge for us, which I think the Committee can think about. Aaron, have I missed anything out?

Aaron Simons

I thought that was very good. I do not have much to add. I thought what was interesting was themes across actually all three challenges, which I guess, as you mentioned, were the fact that there has to be a whole systems approach and this has to go across all stages from the very earliest design stages up to the deployment. This has to be a dynamic and ongoing process for all these challenges. There will not be a moment when we can rubber stamp and say, ‘We have solved this problem’. This is going to be continuous all the way through.

Dame Shirley Pearce

Just one of the things for the later discussion about whether we have got the challenges right, am I right in thinking that there is an overarching challenge of responsibility that manifests itself in a number of different ways? We might come back to this at the end, but if you could be thinking, because it is very important for us to come away with a framework to help us as a Committee make recommendations that are useful to you, so your input would be useful.

Lord Evans

Amy, did you want to say anything?

Amy Austin

I was just going to reinforce what Aaron said about this whole-systems approach. I think that is really key and I think it is quite interesting to start thinking about whether rubber-stamping ethics is actually the most appropriate way of doing things. On first glance we might think, ‘That is actually quite a good way of thinking about it, maybe we should be adding these things into procurement processes’, but thinking of it as more of an ongoing process is really important and something that we have not really touched on or spoken about much. I think that is a key consideration for us going forward.

Lord Evans

There could still be a procurement question.

Amy Austin

Absolutely.

Lord Evans

You would not wish to contract with a company that did not have the wherewithal or the ability to have that continuing dialogue on these sorts of issues during the engagement phase.

Amy Austin

No, I completely agree. I think it is just more about thinking of it as a dynamic between government or the public sector and the person you are procuring the technology from, as both of you were saying. So seeing the procurement of technology more as a service rather than just as a piece of technology that the private sector give to us and we use, without having conversations about how best we do that. I think that is probably quite important.

Lord Evans

Okay. That is very helpful. We are going to move on to the second session, which I suppose is just a continuation of the conversation. I want to start on this question of algorithmic transparency. As a Committee, the standards in public life are very clear on accountability, objectivity and so on, and a lot of that goes round to this question of understanding decisions and being able to articulate that. Depending on who you talk to and depending on the particular form of AI that is being used, there is a perception that it can be quite hard to understand why a particular decision has been reached, but there also appears to be differences of view as to whether that is intrinsic in the nature of some of the AI technologies or whether that is just because nobody has asked for that to be built in at an early stage and, therefore, nobody bothered. I think this is really quite a key part of the conversation. The key question is what form can and should algorithmic transparency take and what does explainability or transparency mean within an AI context? Who would like to kick off on that, because I imagine almost everybody has had an interesting series of engagements on this question?

Simon McDougall

I do not mind starting. I have already plugged the artificial intelligence stuff we are doing with [inaudible] and the Turing Institute, so I will not repeat that. One thing I have picked up along the way is that maybe in the earlier days of this current phase for people engaging in machine learning there was an awful lot of stuff that was characterised as being black box AI and could not be explained, but when challenged and picked away at it was sometimes a little bit of a lazy explanation and sometimes better design could have made it explainable. I think something we need to be challenging ourselves on is whether the lack of transparency and the lack of explainability is a real necessity for the system or whether it is bad design.

That is not to say that sometimes there may be certain instances where you have a deep learning, machine learning model and it is terribly complicated and that is the only way to do it. I am not saying that is not impossible, but sometimes there is a challenge to be made of vendors and people who are building the system.

Sana Khareghani

Just to reiterate, my position on all of this really depends on the situation. It is use case specific again. This is a real big thing and I think the bottom line for this is some of the stuff we have already talked about, which is accountability and other things that are weaved into using these technologies, because ultimately why do we want to know why a decision has been made? Because we want to be able to trust it and we want to be able to say that, ‘I feel comfortable being accountable for it’.

Lord Evans

Is that it? For, as it were, the owner of the system that may be the answer, but for the beneficiary of it or attentively somebody who feels they have not benefitted from it there may be a requirement that they have. This is a bit like being in a *Kojak* movie.

Sana Khareghani

You are right. I would use an example of perhaps when you go to a doctor and you get diagnosed with something. For all intents and purposes, what is happening inside that doctor’s mind and why they have given you your prognosis is a black box. You do not understand it, but you do trust it because of potentially how long they have been a doctor, their credentials and their schooling. It is all these institutions that build up this level of trust in that situation and you are the beneficiary and you may not like it. What you do is you then go and get another black box person to give you their opinion and eventually you put a bunch of stuff together and you decide which is the best situation for you. Ultimately, when we think about this and why the use case is really important is it is a very similar, tangible thing and it comes down to that trust, whether you want to be accountable for the decision or whether or not as a beneficiary you think that is good enough and whether you feel okay with having received it.

Lord Evans

You may want to challenge the decision legally, of course.

Sana Khareghani

You want to challenge the decision.

Lord Evans

In which case, explicability might be quite relevant.

Sana Khareghani

Exactly. This is why that use case becomes so important and why we cannot put a blanket Band-Aid over, ‘Yes, we want all decisions to be explainable’, and that leads you to a branch of artificial intelligence that has been around for a very long time called symbolic AI, where you can backtrack and know exactly what is happening versus artificial intelligence, like machine learning technologies like natural language processing or facial recognition, using deep learning, machine learning and neural networks, which people talk about as black box and non-understandable. That is not necessarily the case and I am sure Turing will tell us about that next week.

It is a whole different ballgame, in terms of where we think things are going, but, again, if we think about the use case and when we are okay with that and when we are not putting a blanket ‘everything must be understandable’ limits the use of these technologies, in terms of the benefits that they could offer us. I do think we really need to hone in on it and ask what it is we want to know and what gives us that confidence, legally, from a beneficiary perspective or from an accountability perspective, to use them.

Alex Lawrence-Archer

We spoke when we were talking about responsibility about the different levels and different types of people involved in that system and I think transparency in a similar way maps onto the different people involved. Transparency will look different for the person or company affected by the decision than for the person who is making the decision in response to a recommendation, the maybe person who has commissioned it and then another person who is reviewing the whole thing.

The second point is transparency needs somewhere to go. We had spoken about legal challenge, but hopefully that is not the first resort. If you are creating systems that are transparent you need to make it clear what the person affected can do and hopefully that is a clear system through which they can appeal that decision and take it to that same authority in the first instance.

A third one, which is related but not really in response to your question, is this: I think there is a higher order transparency problem, in that I do not think we know who in public life is using AI or at least we do not know that systematically. I think that the New York example is quite instructive in this sense, in that their initiative has kind of hit a brick wall because they do not know who is using AI. They have been unable to fulfil their mission precisely for that reason. There is a job to do in figuring out who is using it before you can regulate it.

Marion Oswald

I think the legal framework is actually quite useful in this context to look at, because there are fundamental requirements around duty to give reasons around public sector decision-making, so, as you were saying, we need to think about what the person would need to say in order to fulfil that duty and, therefore, what sort of information is required out of that system to fulfil that duty.

The challenge side of it is also really important, in terms of individual rights and, for somebody who may want to challenge that decision, what information they need. Then there is the audit side of things. What does the organisation need in order to be able to properly audit this tool? I think some of the data that comes out of these experimental systems, I am afraid particularly in academia, is pretty bad. You get some claims of accuracy levels of more than 80% or 90%. I often go and actually read the paper – I know people do not expect people to do that – you find that, actually, there is a huge level of false positives or false negative around the inputs and they have all sorts of other things wrong with them. This Committee maybe could think about how standards could be set about that sort of validity of information, about how a tool is performing and what is acceptable and what is not acceptable and then tell off some academics.

Fiona Butcher

As you say, like a lot of the departments, we have been doing a lot of thinking on this particular topic. Just in terms of language, some of the language we are now beginning to use, which again is based on some of the published research, is about the ‘interpretability’ of the AI. Transparency is making sure that the system is transparent enough that you can understand how it works and that comes down to being able to interrogate the building blocks of the AI. Then the ‘explainability’ is about how it behaves, so how it explains the way in which it has reached some of the conclusions and recommendations or analysis that it is providing to the human decision-maker.

The interpretability comes back to Simon’s point about it depending upon context and who needs to know what as part of that systems approach. If I just take an example of driving a car today. It is almost like having a computer under the bonnet of the car. As a driver, I do not actually need to understand how that computer is working, but I do need some information on my car’s dashboard to tell me when I need to fill up with petrol, how fast I am driving and if there is a fault it may give me a warning light. I can take it into a garage and the person in the garage will plug it into a diagnostic system, which will then give them new information on the performance of the car. Therefore, interpretability depends what you need to know. The transparency and explainability are distinct things. Transparency is about understanding how it works and explainability is about how it behaves, just as a human explain why they have reached a particular conclusion.

Lord Evans

There is an interesting question in the literature around the extent to which humans are in a very good position to explain why they have made a decisions, although there is a perception that they have [inaudible]. Jane, you wanted to come in.

Jane Ramsey

I just was listening with great interest to all of it, but I was just thinking from Sana and you as well that one of the things that occurs to me is what the role of intellectual property is in algorithm transparency, because who owns either the commissioning of it or the provision of it? It is pretty clear around machine learning in medical devices, for example, who will own the provision of it, but not necessarily the commissioning of it. Will that not need to be factored into this debate in terms of machine learning and AI in a way that it is not when it is a direct relationship with a doctor?

I think your analogy was a very good one, but there is something then in between machine learning about who owns, as I say, the commissioning of it, which will normally be a public service. It will be the Department of Health, NHS England or whatever. There will be a person who then provides the machine learning and the new algorithms that are not just in the doctor’s head and it will cost money to produce it, so there will be an additional owner who we have not talked about, which I am sure has presumably developed in defence and developed in medicine up to a point, but for local government and for the police I think it is an interesting point.

Ian O’Gara

Does it come back to a data owner and a data processor?

Simon McDougall

I do not think it does in this case. I think it is much more about a trade secret issue and my mind turns to the ACLU Idaho case in the US, where the Idaho state had bought what they thought was an AI driven piece of software, which was going to allocate Medicare payments to the citizens of Idaho and in some cases cuts those Medicare payments by up to 30%. The ACLU took a court case that ran for something like four years to try to get transparency of why these decisions were taken and the software provider was saying, ‘You cannot see this algorithm because it is a trade secret’. In that case when they finally won their court case four years it was a pretty shoddy database with some very ropey macros in it. It was not AI at all, but the tension there is going to be the same. People are not going to want to disclose these. I do think there are interesting ideas bouncing around as to whether you end up having, say, trusted third parties, who can audit algorithms while maintaining confidentiality, trade secrets and those kinds of things, but this is a very nascent area.

Jane Ramsey

Can I just say it is trade secrets, licensing, patents and copyright? That could be a whole other debate that is not for this table, but that would be interesting, because that is a slightly different thing to trade secrets, is it not?

Simon McDougall

Absolutely. There is now some secret sauce that was not there before.

Lord Evans

Alex, do you want to come in?

Alex Lawrence-Archer

Just very quickly, we have talked a lot about context and that we have different tolerance for opacity in different contexts and the types of decision being made. I think we will need to think about what practical tools we can give to people to assess the context that they are operating in clearly, so that it goes beyond a recommendation to think about the context that you are in. I think that the Canadian Treasury Board Directive model is really useful, in that it gives literally an Excel spreadsheet that allows a person introducing an algorithmic decision-making tool to assess their context and it spits out a score at the end, which tells you how tolerate you should be of lack of transparency in that context. Something like that I think is potentially useful.

Oliver Buckley

Transparency is the basis on which it comes up.

Lord Evans

We have come back a number of times to the importance of context. The concern about context is that if you have varying context and, therefore, different ways in which you need to address this that if fine, so long as you have an organisation that actually sees its responsibilities in that way and is able to do that through with a process with integrity, intellectual clarity and so on. The danger is that if you do it right across the public sector you will probably get very considerable variations in the extent to which organisations are able to go through that in a way that commands a lot of confidence. Some organisations can probably do that brilliantly and there are probably some parts of the public sector that would not find that a very easy thing to do if it is unfamiliar territory to them. I guess that is one of the risk areas here at least.

Sabine Gerdon

To come back also to the Canadian Automated Decision-Making Directive, I think it is a very interesting point that we have not really talked about also about the transparency of the public sector on how they use those technologies. To your point that you made, if an agency is not able to go through those processes and does not feel competent enough it is also then the question of if they need to also make transparent that they are using AI we are then in this context specific go or no-go decision.

Yes, this will slow down adoption of AI, but maybe at this point when you do not have the right skills in your department or the right ways to monitor the algorithm nor the decision-making, then you should also be confident to make a decision of not adopting it, because if you need to be transparent of which decisions you are supporting with algorithms I think there is a lot more public scrutiny on this. It is a fine balance, because, like Simon said, we want to make the best out of the adoption of AI. We want to make sure that the public sector can use it, but in those circumstances we also then need to be confident to say that we are adopting it the right way.

Bethan Charnley

I have just two points. First of all, without risking getting too academic, I think there are a lot of problems with the word ‘transparency’ and using it in a AI context, because I think you need to get to the heart of what you actually want. If it is interpretability or if it is explainability, let us just use those words. If we are talking about transparency, there needs to be a recognition that we are using it as a catch-all term to refer to all these various issues that sit within it and there is a risk that if people do not understand it is used in that context they are going to start jumping to problem statements around the need to fully expose the IP of a company when actually you can achieve explainability potentially without exposing any IP. I think that is really important.

The other thing that I would add is that when we are comparing between different models or machine learning algorithms they are often portrayed on this axis of accuracy versus complexity and it is always portrayed as a kind of trade-off between compromising on accuracy if you want to be able to explain it. Thinking back particularly to a conversation with the Ministry of Justice around their need to understand how an algorithm worked extended far beyond the need to explain it to an end user, but actually to explain or to be able to understand the various different levers that were going into making a decision. For example, it is no use to them necessarily having a really accurate tool that can predict how likely this person is to reoffend if they do not understand how that decision is being made, because all of the various other policy interventions that they need to pull together to ensure that reoffending is less likely means they have to be able to see how that decision is being made. It is not just about the end user, but particularly in public life how the decisions are made and how that informs the other activities that we do.

Lord Evans

We need to consider the issues in a full system sense.

Bethan Charnley

Yes.

Ian O’Gara

Whether it is explainability, interpretability or transparency, from a design perspective if you can bake in early, then you can do it. You can get to these levels of insight. It will depend on what your need is. Is it for auditing purposes? Is it for a legal definition? Again, all these things need to be codified and standardised, but they can definitely be designed from the outside in.

The standards are quite important. They need to be loose enough to be able to allow for and not strangle the development of the technology, the marketplace and the capabilities, but they will be useful. There were some interesting points there earlier about third party auditors, creating frameworks and codes of standards to try to stop the backstreet or under the arches types of people coming through, but that auditability is going to be an important process. We can talk later on this, but there is a correlation with the standards around sustainability and ISO standards that came through in the mid-2000s and it started to move the discussion forward. I think we are in that stage now where it is so nascent. We have this landscape and we are trying to work out what that framework looks like.

Fiona Butcher

Just to pick up the point about the ISO standards, these are under review at the moment in recognition that there is a need to change the existing human-machine interface standards with the advent of this new technology. They are being reviewed at the moment and obviously we are engaged in that discussion. With regard to the human-machine interface, it also raises a point that somebody else made around the skills. There is a real training and education piece here. There are varying opinions about whether this technology will deskill or whether there will be a requirement to upskill. There may be skills around critical-thinking and new skills that need to be trained to ensure that we are making the most of AI systems and so that they are assisting with some of that better decision-making that we are looking for.

Lord Evans

Are people aware of examples of good practice in that sort of area of training for those who are going to be responsible for the totality of the system? It is quite a challenging thing, is it not, if it is not something you have done several times before? I just wondered whether there are examples where this is being done really well.

Ian O’Gara

We have seen a lot of private sector organisations in the banking industry and products and they can obviously see the numbers of Coca-Cola bottles being sold, but trying to explain how, why, where and what was the influence has always been a challenge. We are working with a number of companies and helping execs to get to the point of asking those sorts of questions and that cascades down and creates a change in the culture, because once people have been asked those questions a couple of times and they do not know the answers they then seek out that training. My bugbear in all these things is when someone turns around and says, ‘I am not a technologist’, or, ‘I do not do technology’, and then relies on someone else, but they have 20, 30 or 40 years of experience that is part of that critical thinking process.

Fiona Butcher

Within defence standards there is the human-centred design approach, which, again, goes back to what we have said a number of times about taking the whole systems approach and considering what are the training requirements and when do training interventions need to be put in place as part of the system’s development.

Jimmy Elliot

It is very similar actually. What we are seeing now is that people want us to pull together those different threads and provide a simple package. ‘Can you at the outset give me a baseline on how I will use this system in an ethical? What is the training and development my people need?’ I think we have had more experience from that from clinical drug trails and things like that, where people were very conscious of what they were doing. In parallel at the same time, asking the research and development divisions in a lot of our companies to build more capabilities out of the box for [inaudible] interpretability, so there is more that perhaps we have not done as standards. Yes, there is more you could do in the design phase and build phase, but perhaps it would be easier if we were starting from something that was a little bit better. I think that is a recognition in the industry now that there is a bit more we could do to begin with.

Simon McDougall

I have to say there is a lot of overlap with that kind of training and that thinking and the way senior management have had to evolve over the last 20 or 30 years to understand FISCO management reporting and just data in general. It is an extension rather than new.

Lord Evans

I have had a lot of conversations about cyber-security with boards and it is a very similar issue. It is quite a complicated technical issue, but you can ask questions around it without being a computer scientist and get a valuable conversation around risk aspects.

Sabine Gerdon

I think, especially on that point, what we have learned from a lot of private sector firms that have done that effectively is that they focused on the board level, so educating through the different bureaucratic structures within the firm to ensure that the training is on all levels, as well as multi-disciplinary across disciplines. Basically, I think what this means for the public sector is not only having the technologies, but also having policy professionals as well procurement professionals as part of the training programme.

Sana Khareghani

I just wanted to also say on this point around transparency, understandability, bias, accuracy and all that stuff it is vitally important and a lot more research needs to be done here, just in essence. We need to as a country agree that this is an area that needs a lot of research and that actually puts us in quite a leading thinking position. We have institutions, like the Alan Turing Institute, for example, that are putting all their efforts and all their thinking into safe and ethical AI and thinking about the research that underpins some of the crunchier issues that we have been speaking about.

It is interesting when people say, ‘I just want to see the algorithm’. Actually, if we assumed IP was not even a problem, just looking at Google’s algorithm is not going to tell you whether or not this is biased or understandable. That is not going to help you understand what they are doing, other than pitting you in a situation where you get into a legal battle about whether or not this is somebody’s protected piece of work. That is not the answer, but research does need to start.

Simon McDougall

One of the outcomes of that, in terms of where we are right now, is to accept that this notion as to what we think explainability means and what the general public expects around explainability in different contexts is going to be a moving target for many years. As a society, again going back to the work we are doing with ATI and explaining AI, I do not think we could claim there was a consistent logical response in different contexts. They are still unpicking that, but it is an emotional response.

The other thing to say about that research that ties into this is we had to spend with one of the citizens’ juries two full days teaching these members of the general public about AI before we could get to a stage of having a structured discussion around these particular scenarios. We do not have two full days with the general public out there, so as we go along they are going to be picking up things from *Daily Mail* stories, talking to their friends or a Facebook group around how all this stuff works. We in the public sector are going to have to be reacting to that and trying to make sense out of these different things.

Oliver Buckley

It triggers a thought that I think probably we have given inefficient attention to in all the discussion today, which is that let us also not forget the ways in which the use of these systems can enhance standards in public life too. It does offer the opportunity for improved explainability and better transparency that can be delivered at low cost and quickly in response to queries and questions.

Simon McDougall

Sorry to interrupt. Sana, I was going to say to your original point around the doctors example, if I go to see my GP I only have 10 minutes and they give me a prescription and I go out and I just trust the doctor. If I knew there was a zero resource way of me to get a little more explanation as to what they were thinking when they gave me that and I could press a button and it came out I would love to have that.

Sana Khareghani

I totally agree. I was just trying to draw the parallel.

Simon McDougall

I completely agree.

Sana Khareghani

That comes to Oli’s point, which is let us not forget the flipside of this, which is we can get to a place where we can have that and more. It is these biomarkers in your system that have led me to believe you are going to whatever.

Marion Oswald

I am not sure we can necessarily compare the two systems though, because the medical profession has an underlying system of regulation, compliance, training and long-standing fiduciary duties in the role of the doctor and we do not have that in the AI industry. We do not have that history, those codes and that regulation underlying it. I think we have to be a bit careful when we are comparing humans and new ways of doing things. We do not forget the underlying regulation behind this.

Lord Evans

We are helpfully jumping on to the last section, which is a good thing, although we are going to jump over the malicious bit in the middle I think and we can come back for that. This is going to happen and it is a good thing. We can benefit from it if we get the culture right that encourages this to be used in a responsible way and there are things about it that have proper governance around it, but that does bring you on to the C word, which is the culture word. This will have impact. As people work in different ways and as the expectations change, it will have an impact on cultures. What are people’s views on what the introduction of AI is likely to do culturally to public service? What needs to be in our minds as we think about how this is going to affect people’s working lives and expectations within the public sector.

Oliver Buckley

A starting point, which again comes back to the first discussion about responsibility, is one of the things in some of the public engagement work we have done in the Centre so far that was quite striking is the sense in which people when describing the proliferation of these technologies talk about a general sense that technology is something that is being done to them. Given that public servants are people too, I guess there is the danger that that is a cultural mind-set that needs to be challenged in the public sector. The starting point culturally has to be we are in charge, we are responsible and, therefore, when we think about the culture in which these systems are embedded it needs to be clear from the outset that humans remain primary and these are tools that serve our ends.

Jimmy Elliot

Do not follow the satnav over the cliff.

Sabine Gerdon

We have an opportunity here to shape the strategy to take a strategy approach on making the decisions on where we want to use algorithms or where we want to adopt AI in the public sector to shape this human-algorithmic interaction. There are some statistics that show that if we would fully invest into the doctrine of AI 30% of civil service tasks could be taken on by AI. In the future when we think about this, I think it is very important that we think about what we want and what the opportunities are for AI in the public sector and which parts AI could do very effectively, so that then civil servants can focus on other tasks and ensure that we have a strategy behind what those other tasks are and the more valuable tasks that make us happier in our job and ensure that we have control over those decisions.

Sana Khareghani

I think this is one of the most important questions to ask and this comes back to some of the things we have been saying throughout the day about thinking about this from a systems view. If you think about automating, say, the easy parts of a person’s job so that they can concentrate on the more difficult ones and the ones that need the most creativity etc. and if you think about where this has begun in practice what you have is a person who, if you take their day, it is 100% work. 80% of their work is easy stuff and 20% of their work are difficult questions. If you automate that what you leave the person with are only the difficult questions. Two times out of 10 they were answering difficult questions. I have a pretty happy life. I know come to work and all I do is deal with difficult questions. Yes, we have automated all the easy bits, but have we really taken into account what we have done to this person’s life? Are we really taking out the slack and leaving only creative parts and the crunchy bit? Are we doing that?

I think there is a lot of consideration and from a public standards and public sector perspective, there is that duty of care that we should take in terms of thinking end to end around reskilling, upskilling and lifelong learning. What we are doing to the public sector by introducing automation or decision-making algorithms is one of the most important questions and the one that we should be really leading on and setting an example for the private sector.

Fiona Butcher

One of the advantages that there could be potentially from this technology is, say, within the defence context it takes away the dull, the dirty and the dangerous and it could be protecting people, but, absolutely, we are also very cognisant of the fact that you need to be thinking about the cognitive or mental load that you may be putting somebody under. You could be overloading them with lots of difficult questions or lots more analysis than they have ever had before, or you can under-load them, which is also very dangerous. For example, if you just give someone a watching, supervisory role, in which they are just monitoring and not having to interact and then all of a sudden ask them, to make a really critical decision. It is really important to think about cognitive load within the organisational context of decision-making.

Ian O’Gara

There is then that other gap, which is that it will allow for things that are not being pieced together at the moment. When you look at serious case reviews, there is data sitting in siloes that has not been tapped and that can, going back to your point Oli, save someone’s life. I always go back to Corey Junior Davis, who is the young kid who at 14 was killed when for years he and his parents had put data into Newham and into the police and it just had not been picked up. He ended up involved in gangs and shot dead in a park in east London. I am not saying that AI can magically solve that, but it can certainly help. We should not shy away from the positivity that this can bring. You mentioned the two day training. It maybe would not be a bad thing if we all had two days of training throughout the whole country and if we had a bank holiday for AI or something!

It is getting people to understand that, yes, there are some pitfalls and, yes, there are some dangers, but there are some massive opportunities out there. We can either shy away from it and hide away from this whole process or we can start to act and embrace it as a society. I look at public safety analytics projects around the world and I have to say I am pretty proud to be British. Just the fact that this is in existence here and this debate and discussion is happening with so many members of the public sector is just an incredible thing.

Simon McDougall

I will, in a way, build on both those points and then try to go even rosier about this whole thing. If you talk to lots of public servants over the last 20 or 30 years and maybe longer, many of them feel that they were more engaged with the real world and the public in the past than they are now and different tasks of their job have made them less engaged. This is across the board. AI may be an opportunity to really change the tide on this and help people connect with people more and help people connect with the public more to have more human to human contact, because, again, the routine and the dull is taken away to an extent. Maybe from a private sector point of view you might sit and go, ‘That is great. Our P&L will be better for this,’ but from a public sector point of view we can say it just means there is more time for people to engage with individuals. Flipping it around, yes, AI may have helped a young person. More time with a social worker may also have helped, so how do we get more contact time from this?

Marion Oswald

I suppose that will work, provided that the culture allows the public servant always to think about the individual and not always to be routed down the line of saying, ‘This person has to fall into one of these categories in our AI tool’, because there will be people who fall outside those categories and there needs to be a culture to allow those people to be considered individually and something different happen to them. The categorisation of people into only one or two groups and no provision for allowing the individual to be considered outside the system is something that the UN report only the other day was talking about.

Simon McDougall

[Inaudible] went back to preventing abuse, there is this issue about in particular vulnerable individuals being categorised across the board and the more cross-data sharing you have, etc, the more people get into a bucket and never get out of it.

Lord Evans

How specifically do you think we ought to think about helping to build an ethical culture? Do we need formal professional standards in this area? What sort of level of training and education do you think that we will need in order to make this work ultimately for the public good, which is what ethical standards are about, and to be able to do that context-specific judgment that we have been talking about?

Jimmy Elliot

If you think about ethics more broadly, when you see organisations that are ethical and those that are not so ethical, the ones that tend to come out strongest, yes, will have the standards and the policies and those things are important. Obviously, in our business world as well a lot of people do not read things. The ones that are really strong from an ethical point of view do the simple broad things really well that might sound a bit fluffy, like recognising individuals just for who they are at work and having conversations around culture and ethics routinely, so not perhaps policy driven, but making sure from a leadership perspective people are going out there regularly and having conversations around these things, so it remains front of mind, part of their purpose and why they come to work. If you do that that gives you the biggest lift, but it is not necessarily the easiest thing to embed.

Oliver Buckley

I say this from an organisation called the Centre for Data Ethics, but I worry about the notion that data ethics are somehow different or other.

Lord Evans

Is there a comma in your title? Is it the Centre for Data, Ethics or is it the Centre for Data Ethics?

Oliver Buckley

There is no comma. To your point about the cyber-security conversation, throughout there is this sense that people feel unable to engage with these issues, because they fear a lack of expertise and that somehow there is some kind of mystical, confusing thing. In reality, people understand the values and the ethics associated with their jobs. It is exactly the same. What matters is are these tools working in the service of those values and ethics or are they undermining them? That is what people need to pay attention to, but do not feel that they have to come up with a whole new set of ethics. It is not different.

Fiona Butcher

Just building on that point and thinking about professional standards, a lot of professional standards have ‘Competence’ within their ethical codes of conduct. I think that actually speaks to a number of the points that we have been making about training and education, and making people feel confident that they have the skills to work with this technology. I do wonder whether ‘Competence’ might be an additional standard to think about to ensure that as public servants and the public sector we are competent to procure and use the AI technology at an individual and organisational level.

Lord Evans

To address these sorts of questions.

Fiona Butcher

Competent to work with the AI technology.

Lord Evans

Okay. That was very helpful. Thank you. Let us go back to the one that we jumped over, which is preventing abuse. There is a question really as to whether this sort of technology presents the risk of new sorts of problems that we had not previously thought of or if the worst comes to the worst, we end up in a situation where malicious behaviour could be facilitated more easily. I do not know whether that is the case, but it seems to me quite important to at least think about this. Are there things about this that raise new concerns that we probably ought to be thinking about at an early stage in order to see them off at the pass?

Sana Khareghani

Yes, absolutely.

Lord Evans

We will just make a little list.

Sana Khareghani

For me at least, the biggest worry about it is that in other cases when somebody is doing something bad they are putting various things together and then you can kind of figure out that they are doing bad things. In this situation, you do research, you push things forward and you find an algorithm that can be used in a very good way, but also in a very bad way. What lets me sleep at night is the Centre for Data Ethics and Innovation creating and looking at some of these things. Honestly, for me it is a big deal to think about this double-edged deepness of artificial intelligence technologies. Shoshana’s book, *Surveillance Capitalism*, is really interesting. It creates things we have never thought of or did not think could exist or would exist and now do. I think we need to be thinking about how to stay ahead of this.

Simon McDougall

One thing that particularly worries me in regard to the public sector is behavioural insights and nudging, because done in a benevolent way that is normally seen as a good thing, but in a way what we saw with micro-targeting, Facebook and Cambridge Analytica is still nudging. It is basically manipulating people when they do not know what is going on to do something slightly different. Nudging in some ways is just lying.

Now we are in a world where there is going to be far more information around every citizen and there is a very fine line between a public authority of some kind using all the data they have to encourage the right kind of behaviour in the right kind of way and manipulating either the public at large or groups or whatever in a way that may or may not be to their benefit or others. It is a sliding scale and because it is a slippery slope, to mix my metaphors, it is the kind of thing you could see different public bodies suddenly finding themselves in unpleasant situations and only realising that they have gone too far down the line when there is a *Daily Mail* article or some-such.

Lord Evans

In a sense, the stories about the way in which China is using this for reinforcing desired outcomes are the bad case.

Simon McDougall

In a way, the Chinese are being more honest about it.

Lord Evans

They are saying they are doing it.

Simon McDougall

With social credit score and the like, there is a score and if you do something that the state does not like your score goes down and you lose access to bullet trains and the like. I am talking more about somebody who is going about their day-to-day life and things are changing and they do not realise why it is happening or even notice why it is happening. It might be access to benefits. It might be the information they are receiving online. They do not know things are changing and that is easier now. To Ian’s original point, AI in the broadest sense here is in terms of just the ability to make more decisions with more data more of the time. You can do that on a personal level now in a way in which you would not be able to in the past.

Marion Oswald

Can I give an example of that that might demonstrate? One of the first things that came to the National Statistician’s Ethics Committee was a tool that had been built by a university based on ONS data. What it was doing was generating ethnicity from your name, so you could type your name in and up pops an estimate of your ethnicity. That tool also was estimating religion as well. That tool had been used already by NHS trusts, on the basis that they would assess their patient records and then only send certain medical information to certain category of patient based on that ethnicity generator. That is exactly an example of that that raised lots of legal questions, as well as ethical question about the basis on which you have the right to generate somebody’s sensitive data from their name without their knowledge and then use it to do some sort of nudge.

Sana Khareghani

There are a couple of other really simple examples. Fitbits, for example, do micro-nudging. They nudge you to make you move and that is good. You want people to move. You want to encourage that, but also you can use micro-nudging and targeting to ensure that me and Bethan read completely different things or are shown completely different ads or completely different content, so our individual view of the world becomes very different and that skewed view can lead to many different things. In essence, China is very different, because China is completely open and people in China feel protected rather than invasive. The problem with that is that those technologies get sold to other countries, where they are not quite the same.

Ian O’Gara

Only certain sections of the community feel protected.

Sana Khareghani

Yes, exactly.

Jane Ramsey

It is very interesting, because I have been watching with great interest social media and the anti-vax campaign internationally, which is definitely using Facebook to target people they think are going to be more likely to believe the non-scientific nonsense that is the anti-vax movement. What would you think would be a readily public sector counterpoint to that sort of thing?

Simon McDougall

As in an example of the issues?

Jane Ramsey

You say to Fitbit or you say to Facebook, ‘Where it is serious health misinformation you cannot do it’. Then you are arguing about accuracy.

Sana Khareghani

This is you starting to get into quite a thorny issue, because you start to bring in things like freedom of speech and difference of opinion, because if that were the case and we fully agreed one thing versus another then it is like saying to the media, ‘You should not cover the other side of the story because it is not true’. I am being provocative on purpose. Again, one of the things the Centre or our online harms white paper is looking at in terms of bias and targeting is to say at which point is this not okay and government can intervene or say, ‘We will make the ethical decision. We will not leave that ethical decision to be made by the guy who is running Amazon’?

Jane Ramsey

Is China intervening though? Is China as an international public body intervening in their own space here?

Simon McDougall

The answer is yes. The social care score is the mechanism they are using to try to control public behaviour going forward and that was originally set up as a way via assessing credit, which is primarily financial, initially just to enable greater access to credit by Chinese citizens, but through using non-traditional information sources to form a view on the credit-worthiness of individuals, which itself is a laudable thing, because you do not have credit reference agencies over there. It is evolving and there are mixed stories of how fast and quite how it is evolving, but it is evolving to be a much broader piece. They are taking lots and lots of broadly publically available pieces of information and then deciding whether that makes you a good citizen or not. It is not just you. It is also who you associate with and who you link with etc., so you are encouraged to unfriend people who may have a lower score, because it will increase your score. You end up going from something that may or may not have had an ulterior motive at the start, but definitely had a good ulterior motive all the way through.

A good example recently, which is not an AI example but is a good example of data misuse with local authorities and the police, was our enforcement against the Met Police and the Gangs Matrix. The Gangs Matrix was an in good faith attempt to identify both the perpetrators and the victims of knife crime in London, in order so that you could identify the individuals and hopefully do something with them. In terms of how the data was acquired and managed, it was really badly done, so it was a victim of just shoddy data practice as much as anything. Obviously sometimes with knife crime in London the perpetrators and the victims are sometimes the same person. Then that data was shared with local authorities, including[?] Newham Council, and then some of these individuals were denied access to services. That is not AI. Like I say, that is not AI, but that is not AI just because the AI was not available yet. You can see how AI makes that whole process easier and with less friction in how you get to that bad outcome.

Ian O’Gara

The great thing we have at the moment is the fact that in China you have a very advanced technology and it is very interconnected and the downside for them is that it has huge consequences. The great thing here is that we are not that advanced in many of the systems across the public sector. Facebook obviously and the private sector is. You pay Facebook to start targeting ads at people, but we can talk about that another time. We are in a great position where we have not connected a lot of these systems and we have not deployed AI, so we really can start to create the standards and structures in which to enable these things to happen. How we do that is obviously for many of you folks to decide and then it is for folks like me to actually work out how to implement it.

Fiona Butcher

Can I just add one point with regard to the human-machine teaming concept? Another concern is when you have systems that are going to continue to learn and some of that learning could be through interaction with the user. There is the potential to either maliciously poison the training data or to be mischievous in the way that you are training your system to develop in the future. Again, there are real world examples out there of putting an AI out into the wild and then allowing the general public to train it and it taking on certain aspects that you would not really want because people are just being mischievous with it. Human nature being what it is, this is also a consideration for those working within the public sector, particularly if people perceive that there could be any kind of threat from the system. This is a consideration that might be slightly different to the normal considerations around cyber-security.

Alex Lawrence-Archer

I think Simon mentioned we do have a project in micro-targeting or micro-nudging looking at what point can we say this becomes manipulative or is undermining individuals’ autonomy and a lot of that is ongoing. There are a lot of factors involved. One is the insights on which that micro-targeting is based. It might be an insight people in general or people like you use specifically or how you specifically are feeling right now, for example. That is probably a sliding scale of unacceptability. It might be about the end and we have talked about how, even if you have a positive end, then that still does not make certain types of micro-targeting technology acceptable. We can definitely give more information on that project as it continues.

Back to the wording, I do not know how advised the use of the word ‘malicious’ was in the question, in that it seems to be more likely to be concerned with not explicitly malicious abuse, but just where different bits of the public sector are pushing the boundaries, because they are there looking to achieve a goal.

Lord Evans

Ill-judged perhaps rather than malicious.

Alex Lawrence-Archer

Using facial recognition is not malicious. They are pushing the boundaries maybe beyond what we think is acceptable. We are likely to be [inaudible].

Simon McDougall

Invariably with machine learning it is just inappropriate targets. You have set up an algorithm to try to give you one outcome and the computer is doing the best it can to deliver you that outcome and that outcome is not correctly phrased or is not optimal. It is trying to do the wrong thing, so it comes out as malicious. To the point around use of AI in terms of engagement, say, with YouTube, Instagram and the like, those algorithms, which have often driven young people to end up consuming quite extreme images, are there purely because if you have been served five videos of the same thing on YouTube they want to serve you something slightly more interesting just to maintain engagement, because they want to give you adverts. There is no malice in that whole thing at all, but it means that a young person can go from fitness images to dieting images to a pro-anorexia website and it is all being done because that is how you keep them interested and you sell more ads.

Jimmy Elliot

I think malice is quite interesting to put in there though, because there are probably three different categories of behaviours that we are looking at here. There are the unintended consequences, the nudging and the influencing and then the intentional. It is a bit like in criminal law with gross negligence, recklessness and intent. There is a subset there for the intentional and the malicious, where we have to perhaps think about there is a balance. The more that you let people know transparency or interpretability, the more they will understand how you are making decisions and the more they can game the system. A key control in your thinking is that balance between the two.

Lord Evans

I am going to start to draw the roundtable to a conclusion and in a moment I am going to invite Shirley to sum up where we got to in that bit of discussion. Then we have the closing wrap-up, which is really, ‘Do we think this has any implications for standards, whether we have the right challenges and whether the principles are still in good shape at the end of the conversation?’ In case people came with something burning that they were desperate to say and have not had the opportunity, shall I just poll around the table? If anybody wants to say anything they have not said or recant on something they said that was heretical or something, here is your moment. Recantation or addition?

Simon McDougall

I recant everything I have been going on about for the last two hours. Some of you know that the ICA was developing an audit framework for AI and Ruben, who is driving that, is at one of the roundtables, so I will not go on about that. One of the things that we are very conscious of there is I think that in a way AI will end up being quite similar to cyber-security for us, whereby while we talk about lots of exotic cyber-security threats a lot of the time, the vast majority of cases we enforce against are bog standard breaches and it is social engineering, buffer overflow or unpatched servers. It is stuff that is entirely avoidable in this day and age, but still it happens. It happens across the board, but it happens a lot in the public sector. I think with AI we should focus on the new and exotic and lots of stuff within AI is new and exotic, but I have a feeling that in 10 years’ time or 20 years’ time we will be enforcing on stuff that is entirely avoidable and we could fix now. In terms of the challenges for the public sector, I think that is one thing to keep in mind. There is going to be a lot of hygiene factors and basic stuff we should be focusing on getting right, as well as the interesting, out there stuff.

Lord Evans

Anybody else?

Bethan Charnley

Just one final point from me, I would encourage you to continue with this work to also engage with the practitioner side of things as well, because something that has really struck me in the work we have done in GDS is the often disconnect between the policy world of AI and the technical world, where nobody really identifies as an AI practitioner. You have data scientists, you have data engineers and you have product managers. The reality of doing AI is often quite different and quite disconnected from how we talk about it in a policy world. As you go forward, you are looking to implement things that land in a context and understanding that reality is quite important for that as well.

Lord Evans

That is very helpful and we absolutely take that on board. That is very helpful. Okay. Thank you. I will hand over to Shirley.

Dame Shirley Pearce

We started off with rhythmic transparency and I think that interpreted that to mean what we mean is trust. What is it that we want to know in order to get trust in the system we are talking about? That relates to explainability and interpretability and perhaps we are being a bit over-simplistic in talking about it as transparency. The question is what do we need to know in order to get trust and then we broadened the discussion away from the algorithms to the whole transparency about AI. I got the impression that there is a concern that we are not quite as clear about where AI is being used and whether the public are as confident as they ought to be about where it is being used and how it is being used. We have a duty to give reasons for our public service decisions and if that is one of the areas that we are not being clear about we ought to improve that.

We talked about the context influencing the level of transparency and all the way through we came back to this context business. I think that is going to be quite a challenge for how we articulate some of the things we wanted to say about standards, because different people want different things at different times and in different places. There was a strong message that transparency should not be thought about as something you think about at the end stage. Like many other things, it needs baking in early. You need to have a way of designing and a culture of design in which transparency, interpretability and explicability are built into the beginning, a bit like with all research projects you should not think about statistics at the end. You should think about statistics and how you are going to analyse right at the very beginning. I thought that was quite important.

We then talked about training and culture. I think this is really interesting, because I got the impression that people were saying, at least for senior management anyway, that this is a development of an educational process that has been going on for some time and we should not think that we have to suddenly rush in and give AI training. This is an extension. I got the impression that maybe we should be thinking that nobody is going to be AI-free or technology-free. Everyone needs to have some kind of understanding of their responsibilities and the different kinds of AI.

It made me think actually of the times when computers came on everybody’s desk and for quite a long time we used to have armies of people who would come round and solve your software problem for you, because none of us knew how to do it. It is a bit like that at the moment. AI is this right that you do not really know what to do with, so you expect there to be some tech person to come and sort out for you, but actually in five years’ time we will all be sorting it ourselves if we have got the education right, because we will have the right skills to know how to solve the problems that we are responsible for.

I felt really proud for the UK about what was being said about it being great that we are dealing with this now. I had not really appreciated that other countries are not quite doing that and if that is the case, please do let us reaffirm that and we will be really proud of what we are doing and it is great that you are involved in it. The scale of the challenge of coming up with some answers to guide the development and to guide the operation really became quite clear and it is quite extreme from some of the risks of this all not going right.

I absolutely take the point that we should not let it feel as if the AI is now in control. Humans are still in control and what we [inaudible] in the standards that we are setting must strengthen that position. I do think that there is quite a concern among the general public, because this is all so new, that they are not in control anymore. That was particularly clear when you talk about the impact on jobs. People’s jobs are changing without them being in control of it. Although we should be creating a culture where we are in control of the process, it does not feel like that everywhere and I think that is something for us to think about.

We were also having a very important discussion about living the ethics and not just having the code. Like ethics and good leadership anywhere, it is not going to be about what you write down on paper; it is what we see people do. I felt quite frightened by the discussion about abuse. I think there are some really big unintended consequences. I was not convinced that we should take ‘malicious’ out completely. No-one would think a doctor would try to kill someone, but they do sometimes, so the whole business about maliciousness should not be lost. I wonder whether we should be doing more in skills or what is happening in schools to help people feel that this is part of something that they are growing up, like their iPads. Perhaps there is something for us to think about among a much broader education dimension than just people who are responsible for implementing AI in public services on the ground. Amy, were you taking notes this time?

Amy Austin

Yes, indeed.

Dame Shirley Pearce

Have I missed something?

Amy Austin

No, I think you have done a really great job. What is also really interesting was Fiona’s point on competence and making sure that people feel confident that they have the skills to engage with the technology, because I think that at this point if somebody said to Aaron and myself even, who are quite abreast of these issues, ‘You have to use AI to help you make the decisions that you make at work’, I think that would be something I would struggle with or not really quite know how to do. Like you say, it is that whole systems approach and to me, part of that is making sure that everybody at every level feels confident and comfortable and is competent to use the technology. That is pretty much it.

Aaron Simon

Yes, I have nothing to add.

Dame Shirley Pearce

What we wanted to do at the end really was to get your reactions to the challenges that we have presented. Are they the right ones? Is something missing? Have we got too many? Are they subsets of each other? This will be a framework that we use to tell the word what we think needs to be done. The framework is quite important and your input in will be much appreciated.

Ian O’Gara

For me, it would be useful to go into it with a recognition that you are not going to get the right answer, to allow for a constant re-evaluation of the process as you move forward and as you test and learn the process with policymakers, data scientists and the public and to accept that this a journey that we are going to have to go on.

Lord Evans

There is an interesting point there, because we had a lot of conversation in the Committee about whether now is the right time to do this or whether we would be better coming back when things are a bit more mature. I think the general feeling from people we spoke to was it is probably more helpful to get involved now, but that also means that this is going to be the final word, whereby we can *ex cathedra* state what the position is, because it is going to be changing. It is a good point.

Fiona Butcher

I do wonder whether it is either a point within the Standards Challenge six or a separate, Standard around the need to ensure the competence and skills required by public bodies and those who are public servants. It is not so much about understanding the technology, but knowing how to work with it effectively. It is some of those ‘critical thinking’ skills that will be needed perhaps far more.

Dame Shirley Pearce

That is a current gap. As a principle or standard though, it is something our report ought to address I think is what you are saying. Does it impact on the standards that we think we should be setting in this area?

Fiona Butcher

I do wonder whether it is a Standard that we need to be explicit about; that we will ensure that our public bodies and public servants are appropriately skilled and competent to work with AI.

Amy Austin

I suppose the point is that in order to uphold our standards, these things need to be in place. You cannot uphold standards if you do not have the skills to do so.

Jimmy Elliot

Is the question selflessness and those principles?

Dame Shirley Pearce

That is the second part of this, really. We currently have the Nolan Principles. Are those principles appropriate and will they be appropriate for digitally harnessed and AI partnered world? Should we be thinking about whether there are other principles that ought to be added in? Some of what I am hearing today is that it is really not that different. It is just how you think about it that is different, so we should be thinking about how we apply these principles in an AI world rather than making up new principles. This is the moment when we should be asking ourselves that question and your views on it. In selflessness, integrity, objectivity, accountability, openness, honesty and leadership is something really missing that AI is going to trip us all up on or is it just the way we articulate them in an AI world?

Jimmy Elliot

They are very similar actually to when you look at other countries and what they are doing in AI. They are looking at very similar things. Generally, they are very strong. As somebody a bit external, perhaps I do not read them in the same way that you might read them, so the question for me would be the way they are articulated do you think it draws enough the human autonomy principles and the individual? You need to think about them. Arguably, it is about how they are articulated.

Simon McDougall

That ties into an observation I have on the six areas and, again, Ruben can talk to this more in the next roundtable. We put out a framework for auditing AI, which has a whole list of things. Looking at and comparing it to the six, there is an awful lot of crossover. The one thing we have as a separate area is exercise of rights and obviously for a privacy regulator that is quite an important thing to us anyway, but the one thing that is not in that list of six is enabling individuals to exercise the rights that they enjoy. Everything else is pulling in that direction.

Dame Shirley Pearce

That is an important point. Are there any other reflections, either on the way we framed the challenges or the appropriateness of Nolan? Nolan lives by the sounds of it. Good. Okay. Thank you.

Lord Evans

Shirley, thank you so much. A little public service announcement, which is part of the reason there were so many police cars whizzing about, is there has just been a bomb scare at Downing Street, but I am pleased to say that it has just been cleared and it was presumably some tourist leaving their shopping bag there or something. All sorts of things have been going on apparently and little robots have been running up and down having a look at it, but they have all gone back to their kennels.

I think it just remains for me to say thank you so much for your presence and your engagement. It has from our point of view absolutely met our ambition, which was to get some of these issues out and discussed in a way that illuminates our thinking process. As we mentioned, the text of this will be cleared with you and then will go on the website. If your organisations would like to make any other formal submission, our submission process is still open until 12 July, so please feel free to do that. We absolutely encourage you to do that. If there are other people we ought to be talking to but are not, then please feel free to let Amy or Aaron know. I am sure we will be coming back to some or all of you over the coming months to check that we are still heading in the right direction, but it has been extremely helpful from our point of view and thank you very much indeed. Thank you.

**This Transcript was produced by Ubiqus UK  +44 (0) 20 7269 0370**

<http://www.ubiqus.co.uk> / [infouk@ubiqus.com](mailto:infouk@ubiqus.com)