



HOUSE OF COMMONS

Science and Technology Committee

Oral evidence: Governance of Artificial Intelligence (AI), HC 945

Wednesday 25 January 2023

Ordered by the House of Commons to be published on 25 January 2023.

[Watch the meeting](#)

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

Questions 1-96

Witnesses

I: Professor Michael Osborne, Professor of Machine Learning and co-founder, University of Oxford and Mind Foundry, and Michael Cohen, DPhil Candidate in Engineering Science, University of Oxford.

II: Katherine Holden, Head of Data Analytics, AI and Digital Identity, techUK, and Dr Manish Patel, CEO, Jiva.ai.

Examination of Witnesses

Witnesses: Professor Michael Osborne and Michael Cohen.

Chair: We are very pleased this morning to be beginning the public sessions of our new inquiry into the governance of artificial intelligence. At the beginning of the inquiry, Members who have a relevant interest are asked to declare it. I think Stephen Metcalfe has one to declare.

Stephen Metcalfe: Indeed; thank you, Chair. I need to declare that I am the co-chair of the all-party parliamentary group on artificial intelligence.

Q1 **Chair:** We have no other declarations, so we can proceed.

This is obviously a hugely important area for society, and indeed for the world. In beginning the inquiry, I—as many others did—turned to ChatGPT to see what they might suggest, and in answer to the invitation to write three questions for a UK Select Committee inquiry into regulating AI, within two seconds, the following three questions came up: one, how can we ensure that the development and deployment of AI is ethically responsible and transparent? Two, what measures can be taken to mitigate the potential negative impacts of AI on employment and on the economy? Three, how can we strike a balance between promoting innovation and protecting the public from potential harms caused by AI?

Those questions have a very uncanny resemblance to the questions we will be pursuing. Not at all defensively, our Clerks responsible for this noted that it was highly likely that ChatGPT had accessed the terms of reference for our inquiry and returned them to us, but we will be discovering during this very important inquiry some questions about current and future deployment, including those three questions and some others.

Let me introduce our first panel of witnesses. I am very pleased to welcome, joining us from the far east, Professor Michael Osborne, who is professor of machine learning and director of the Centre for Doctoral Training in Autonomous Intelligent Machines and Systems at the University of Oxford, and is co-founder of the firm Mind Foundry. Thank you very much indeed for joining us, Professor Osborne. With Professor Osborne, in the room, we have Michael Cohen, who is a researcher in engineering science at the University of Oxford and specialises in this subject matter.

Perhaps I can start with a question to Professor Osborne—our first question of this inquiry, which is a foundational one. How would you define artificial intelligence?

Professor Osborne: Of course, this is an important question, and the definition that I found in the Government's proposal from July last year seemed about right to me. There, it was focusing on adaptiveness and autonomy, which are two of the core characteristics of AI that distinguish



it from previous technologies, but the other thing I saw there was a recognition that the consideration of “what is AI?” needs to be taken on a case-by-case basis. You need to deeply delve into each application to consider the extent to which it is truly AI.

That is, any definition needs to be flexible, and in particular to guard against people ignoring any sort of legislation about AI by putting a human into the loop. You do not just want to have a human dummy rubber-stamping decisions made by an AI, thereby getting around any rules that might be put into place. But by and large, I think, the definition that has been provided seems pretty good.

- Q2 **Chair:** Thank you; we will come into that human-machine interaction in some detail. The fact, Professor Osborne, that you mention approvingly the UK Government’s definition points to the fact that, as I understand it, there is not a universally agreed global definition of artificial intelligence. Why is that?

Professor Osborne: It is because the technology is quite diverse. AI as a field encompasses many different types of algorithms for many different purposes. Different applications can look very different from one another. Another part of the challenge is that the thing we are trying to do—intelligence—is itself poorly understood. It is hard to define what intelligence is, even in human beings and animals, let alone in these poorly understood constructions of our own. For all those reasons, we have to make do with these rough-and-ready definitions, such as AI being defined by learning, adaptivity and some degree of autonomy—in other words, it is able to take actions on our behalf.

- Q3 **Chair:** Thank you. Does the lack of a crisp, standard international definition pose any challenges for the governance of AI?

Professor Osborne: Very much, and it points to this need to consider applications on a case-by-case basis. I don’t think any unifying definition will be arrived at, so there has to be some means of considering each individual case on its merits and governing it appropriately.

Chair: I see. That is obviously a very important steer. This first session is something of an introduction to the terrain, and so your steer is that the governance should be based on applications, rather than a universal conception of AI. Thank you very much indeed.

- Q4 **Graham Stringer:** I find it very difficult to think about AI, Professor. It has always been difficult to predict how different technologies will develop. Famously, IBM, I think, thought the world would need only one large computer. They were wrong. Is it even more difficult because of the autonomy of AI to predict which direction it will go in?

Professor Osborne: It is correct to say that predicting the future of AI is difficult, as has been proven in various forecasting activities held by leading world forecasters who have not yet managed to perfectly predict how the technology will develop. The difficulty is linked to the complexity of the technology and the fact that it is poorly understood. Even the aims



of the technology are not well understood. We have seen repeated cases of the technology vastly exceeding what we had reasonably expected to be done. Most recently, large language models—ChatGPT is just one example—have massively improved on what people thought would be possible in short order.

- Q5 **Graham Stringer:** Having said how difficult it is—you gave a couple of examples—can you make a stab at telling us which way you think AI will go and whether we should be frightened of it or not?

Professor Osborne: It is fair to say that AI will continue to develop. We can expect, if nothing else, continued change. At present, we are seeing enormous progress—in particular, in large language models capable of generating text that could have been produced by a human, and in algorithms able to convert text into images. In that respect, we are likely to see continued advances in generative AI—AI that is able to use tech in the place of a human being.

It is fair to say that we are still only scratching the surface of what is possible with AI. One thing that distinguishes AI from other highly hyped technologies like blockchain and quantum is that it is actually proven. People have realised great value from it today, particularly big tech. We also know that there are many aspects of AI that have not yet been brought to market. There are many capabilities that have impressed in demos, like ChatGPT, but have not yet been realised in economically valuable applications. Predicting the future is a mug's game, but all I can say is that we are going to see continued rapid change in AI for some time to come.

- Q6 **Graham Stringer:** You sort of answered this question in your introductory remarks when you said that you thought every application should be considered on its own merits. In terms of any national or international regulatory system for AI, is it possible to have a universal regulatory system? Do you think different countries can have different regulatory systems?

Professor Osborne: I think it is certainly possible to have some overriding principles informing how AI is governed, and in an ideal world, those would be harmonised among different nations because the problems that we face are very similar.

Some of those principles would directly address some of the harms that AI might pose, many of which have been discussed in depth, but I shall reiterate. Of course, AI poses challenges through its introduction of bias. AI poses threats to privacy. AI is a dual-use technology with military applications that need to be appropriately considered. We also have destabilising influences from AI creeping into our society. In the immediate future, we are seeing destabilisation via social media with AI driving the feeds that we see on Twitter and Facebook. In the not-too-distant future, we can reasonably expect economic disruption, including much churn in labour markets, as tasks and occupations become more



automated. Focusing on the harms is one way to begin to consider overarching frameworks for governing AI.

- Q7 **Graham Stringer:** I have two questions. You mentioned some of the difficult areas to apply AI. There was a debate last week in the House of Commons, and this Committee has looked at the evidence base that the Parole Board used for letting people out. Do you think that AI can help in applications such as whether or not prisoners should get early parole, or do you think that's much too dangerous a precedent?

Professor Osborne: That's a really interesting question because AI is being used for such decisions already. AI is making hiring and firing decisions, so it is not unreasonable to think that it might be able to play some role in awarding parole, but much caution would need to be taken in any such application. AI is ultimately taking decisions in a very different way from the way humans do, and you would want to make sure that the decisions it was taking were supervised by a human. You would want to make sure that a human was making sure that the algorithm was not biased in any sort of illegal way. You would want to be making sure that there was some explanatory framework for the decisions that were made— *[Interruption.]*

Tracey Crouch: Never mind AI.

Chair: If you can still hear us, Professor Osborne, you have frozen on our screen.

Graham Stringer: Shall I put my last question to Mr Cohen?

Chair: Yes, indeed.

- Q8 **Graham Stringer:** It could have gone to either of them. Has the present development of AI made the Turing test obsolete?

Chair: Professor Osborne, that question was directed to Michael Cohen as you were frozen momentarily on screen.

Graham Stringer: Or both of you, possibly.

Chair: Or both, but we will start with Mr Cohen, then the professor.

Michael Cohen: No, I wouldn't say so. I think that if you get into really—well, I suppose you might think about having to refine the Turing test a bit, so there might be people with certain expertise that you could distinguish from an AI at a point in time later than people without that expertise. Maybe the Turing test would be a bit simplistic in assuming that every person is more or less the same.

- Q9 **Katherine Fletcher:** Would it be possible, for the biologist in this session, to define the Turing test?

Chair: Yes, perhaps we should do that.



Michael Cohen: The idea behind the Turing test is that you can know if a computer programme has achieved human-level intelligence if you cannot distinguish it from a human.

Katherine Fletcher: Like a double-blind trial.

Michael Cohen: Yes. Of course, not all humans are the same; they have different levels of expertise, and it could be that there are some humans you could distinguish it from but not others. I suppose you might need to refine it in that way, but I think it is still generally a decent test.

Professor Osborne: Michael is absolutely right, but I will add some detail. First, the algorithms that have passed the Turing test to date tend to do things like mimic being someone speaking in a second language—a child, for instance. This is what Michael was getting at, in that to use the test to more meaningfully distinguish human from AI, we might want to try to identify humans communicating more sophisticated content.

It should also be said that AI today is very far from human-level intelligence, and even ChatGPT and similar large language models have quite problematic blind spots in their reasoning. To take one example, if you ask ChatGPT how you might join two pieces of paper together using only a pair of scissors and a band-aid, a plaster, the large language model—the AI—will be unable to concoct a way to do so, whereas almost any human child would immediately recognise that you can use the plaster to stick the paper together. The use of a plaster for adhering paper together is not something that is in the training set of the AI, and hence it does not understand that that is a possibility. These AIs still have really significant gaps in their understanding of the complexities of the real world. None the less, they might be able to pass a Turing test if the test itself is sufficiently narrow.

Graham Stringer: Thank you.

Q10 **Dawn Butler:** Professor Osborne, thank you for your evidence today; it is fascinating. You mentioned bias in technology. I wondered if you could drill down on that for the Committee and help us understand the different types of bias in machine learning.

Professor Osborne: Bias is a huge issue for machine learning—let's make no mistake about it. Some people present bias as purely being a result of garbage in, garbage out, and of course, that is a factor. If you train a machine-learning algorithm on a dataset that is already biased, the machine will faithfully reproduce that bias. Famously, in 2018, Amazon was forced to admit publicly that an algorithm it was using for hiring was fundamentally sexist—that is, it was privileging applicants who identified as male over those who identified as female, even when the CVs were otherwise identical.

Unfortunately, the problems of bias do not stop there. You cannot simply train up an AI on a perfectly unbiased algorithm and expect it to do the right thing, because fundamentally, AI is not always meeting the right goals; it is meeting the goals that we say, not the goals that we want. I



will give a non-AI example. Back in 1908, in Paris, there was a dog that was lauded as a hero for rescuing children from the Seine. A child had fallen in, and the dog had rescued it from the river and been rewarded with a piece of steak. It happened again, and the dog got another piece of steak, but it then became a recurring pattern—this dog seemed to be rescuing children every other day. When the case was investigated, it was discovered that the dog was actually patrolling the banks of the Seine and pushing kids in so that it could rescue them and get the steak. This is a real problem with AI in general. We have to be very careful about the goals that we set it, because it is going to ruthlessly attack those goals, even if they are not actually the goals that we meant.

- Q11 **Dawn Butler:** Absolutely. I am with Graham: I am very worried and nervous about AI for that reason. How do we stop people believing that AI is the future and therefore putting all their confidence in it, which will make it difficult to challenge the AI system?

Professor Osborne: Again, that is a huge problem. Do you know the term “bionic duckweed”? If not, and for anyone else who does not, it refers to a common pattern of thought that says, “Well, we don’t need to solve the challenge of electrifying our transportation now. We don’t need to come up with new ways of getting renewable energy into transport, because somewhere down the line someone will come up with a way of biologically engineering a form of weed that we will be able to grow in enormous proportions. We will then be able to use that weed as a biofuel, and all our problems will be solved. There is no need to do anything now, because in a couple of decades we will have this magical technological solution that will address all our issues today.”

It is a thought pattern that occurs very often in AI. People are looking to AI as a kind of magical solution that, somewhere along the line, will solve various problems that we currently face. I am absolutely with you that that is not the way that we should be thinking about AI. If we have problems, we should find ways to solve them today, and not look to AI as some kind of silver bullet solution that is coming down the line, particularly because AI today is so compromised by so many issues. At present, we have no idea how to address most of them.

- Q12 **Dawn Butler:** Thank you, Professor. I do not know if Michael wants to come in too, but I will throw in the police and their use of AI, which is quite prevalent in America at the moment. There are lots of cases, but there was a recent case of a black man who was in jail for a week, accused of robbery in a place he had never been to in his life. How do we mitigate that and stop it from happening?

Michael Cohen: That is a very good question. Professor Osborne was starting to talk about where this behaviour can arise from, and the first thing he mentioned was garbage in, garbage out data—if the data is biased, the predictions will be too. I echo what he said, because that is not necessarily the only problem.



One way to understand human racism, or bias in general, is using the quickest, roughest heuristic you can come up with—the one that strikes you first—and running miles with it. If you see that there are not many female CEOs in the world, the first thing you might conclude is, “Well, that is not the sort of thing that women do.” One of the things we use our intelligence for is to look closer and see that that is not really what is going on at all. If you have an AI that is looking for the simplest explanation—the quickest explanation it can find; maybe its model is not rich enough to even be able to express what is really going on—you will end up with something that cannot even conceive of its own limitations, and by necessity relies on simplistic and ultimately incorrect explanations. I have not looked at what exact algorithms the police are using, but I suspect that that is one issue involved.

As for solutions, people can try conditioning predictions on certain things, and there are various technical proposals for how you might enforce certain indifferences and make it discount that. Typically, to the extent that the thing you are conditioning is poorly measured, there will be some leakage, so that is difficult.

Q13 Dawn Butler: Can I ask you both one final question? Are all datasets biased? Do all datasets have a bias in them?

Professor Osborne: If I may, I will briefly return to your question about policing. The company I co-founded, Mind Foundry, has actually been working with the Scottish police force in bringing AI to work there. But the way that we have been doing that illustrates a general point that I want to make, which is that there are right and wrong ways to deploy AI. The kind of use of AI for policing that you were describing is one that I would describe as a no-go area; I would not use AI for that at all. Instead, we recognise that there are many places to help police that involve just the automation of much lower-level tasks. For instance, we are using AI to help process case reports and to do named entity recognition—to pick out the relevant pieces of text to help unify large databases. We should always keep in mind that there are ways for AI to help that do not necessarily go anywhere near the things that we do not ever want AI to do.

To your question about whether all data is biased, I think that is absolutely right. Fundamentally, data is gathered by human beings for a particular purpose, and data always encodes some decisions. Data is a result of a decision-making process about what is worth measuring and what is not, and it involves some sorts of boundaries being drawn in the world—some sort of categorisation. None of those categorisations can be viewed as completely neutral; it is the result of some assumptions on that human’s part. It is an illusion to think that data is neutral and objective. It always encodes some sorts of biases in the form of assumptions, and they always need to be questioned and examined before any AI is put to work on that data.

Dawn Butler: Fantastic.

Q14 Chair: Before we go to Tracey Crouch and then Aaron Bell, I have a



HOUSE OF COMMONS

follow-up question for Professor Osborne. You mentioned machine learning a few moments ago. What is the difference between machine learning and artificial intelligence? Are they interchangeable?

Professor Osborne: Sorry, I conflated those two terms. They are distinct, but not very. AI is the broader field. AI is a superset of machine learning, but the other areas of AI that were traditionally distinct from machine learning have, in the last decade, more or less been taken over by machine learning. Traditionally, computer vision, which is the processing of images, was seen as something distinct from machine learning but is now almost entirely a field of machine learning itself. Machine learning is the use of statistical and other algorithmic techniques to solve learning problems. But by and large, today you can consider the two—AI and machine learning—as the same thing.

Q15 **Chair:** I see. Thank you very much indeed.

I have a question for Mr Cohen. We have already surfaced some of the problems and worries about the application of AI, but your submission in evidence to us referred to the economic value and positive aspects offered by different types of AI. Could you summarise or reflect on some of the areas that offer significant benefits to society and to the economy?

Michael Cohen: As Professor Osborne mentioned earlier, ChatGPT is a long way from being able to imitate humans extremely well. But in the future, if we have really high-quality imitations of humans, a lot of their economic output could be produced much more cheaply. This is very close to a problem but, looked at on its own, it is a great possibility. You could imagine the economy growing unfathomably. I will get into issues with that, if you like.

Q16 **Chair:** We will do that in due course. You can do things more cheaply in terms of resources—that is clearly one aspect. What are the other benefits?

Michael Cohen: Producing things cheaply is a pretty broad category of benefits. I could get into the specifics of things you could produce more cheaply, but that is about all we can ask for from technology.

Chair: Let me bring Stephen in, and then I will ask the same question to Professor Osborne.

Q17 **Stephen Metcalfe:** This technology has been around for quite a while. It is improving, growing and developing all the time. I started the all-party parliamentary group on artificial intelligence in 2016 because of the phrases that were being bandied around and because of the potential impact on people—in my case, constituents—being replaced by AI in all its various guises. And yet six, seven years on, we have not really seen any impact. What AI has done is augment the way people work. Can you give any specific examples of where AI has been deployed to replace people, as opposed to augmenting what they do?

Michael Cohen: No, I cannot. The best I can give you is an example of the economic output of horses after the combustion engine was



developed. That would be an instance of a technology really just replacing the work of some other thing.

Q18 **Stephen Metcalfe:** That is one specific area, isn't it? That was in transport.

Michael Cohen: Right, but that's all horses could do, really.

Q19 **Stephen Metcalfe:** We're a bit more flexible.

Michael Cohen: Right, and I think that is why you haven't really seen that with us yet, because AI is not at the level where it can do what we do. I do not think this is the sort of thing where you will see gradual encroachment so much.

Q20 **Chair:** We thought we had lost Professor Osborne, but it turned out that the lights went out, so it was a very low-tech problem. Professor Osborne, we have started surfacing some of the benefits. Doing things more cheaply in resource terms is one benefit; are there any others you would like to add at this early stage of our deliberations?

Professor Osborne: Mr Cohen is right, of course, that being able to do tasks more cheaply than a human offers up enormous potential, but beyond that, AI does not suffer from some of the problems that afflict human labour. For instance, AI can be far more vigilant. AI can work 24/7, and it does not get distracted by kids in the background or anything. AI is scalable to a degree that humans are not. You can spin up as many instances of an AI as compute as you can. For all those reasons, AI can do things that humans cannot.

AI can also operate in extreme environments. You can have AI on satellites, as indeed we are doing. There are many potential improvements from AI, just viewed as a replacement for human labour, but as you have correctly said today, AI is better thought of as an augmentation of human labour—as a collaborator with humans. In that respect, it is already having an enormous impact. AI might not be replacing wholesale occupations, although if I can give one fun example, in 2013 we predicted that fashion models were highly automatable. We predicted that they had a 98% probability of automatability and we were laughed at. But now, of course, there are firms producing digital models, with the aid of computer graphics software, that can pose in whichever clothes you want, to produce digital images that you can put up on your social media profile, for actual fees from fashion brands—so fashion models are directly being automated through the use of AI technologies.

Q21 **Chair:** On this question of the current role of augmenting the work that humans do, is that not just a current and temporary restriction? Will it not be the case, either now or shortly—for example, in the diagnosis of medical conditions—the capacity to analyse that an AI package might have is simply beyond the computational and cognitive powers of the most eminent physician, in the same way that some of these games players are now regularly beaten by machines? Aren't we going to get beyond helping people to actually supplanting them, because AI can do



things better?

Professor Osborne: I think your first point is absolutely right. There is no law that says that AI will always be limited. There is every reason to expect that, even this century, we might see AI that is at least as capable as human beings, and very likely far, far more capable than any human being is today. That is a prospect but, at least to my mind, it is not an immediate one.

You are absolutely right that today AI is better thought of as augmentation. Taking your example of game playing, it is important to recognise that a game is an environment that could not have been better designed for AI to perform in, in that a game is very crisp and clear. The rules of the game are provided explicitly to the algorithm. It is usually easy to gather an enormous dataset of games played, so the AI can learn the moves, and usually, there is perfect information. Usually, you can see the board state without any of the mess, lack of structure and heterogeneity that makes real-world decision making—real work—much more difficult. We should not over-index on the success of AI in games, because games are uniquely suited to AI.

Q22 **Chair:** And they don't have any of the ethical questions that Dawn pointed to, for example.

Professor Osborne: That is exactly right. In your example of medical diagnosis, I think those ethical questions are very salient. In fact, I do not think that we should be using AI for medical diagnosis immediately. Notably, despite really optimistic forecasts—such as that from one of the pioneers of deep learning, Geoff Hinton, back in 2016—that all radiologists would be out of a job within six years, here we are seven years later and that is not true. There has not been this big uptake of AI technologies, even in those areas of medical diagnosis where computer vision is perhaps best suited.

We wrote a report on how AI could contribute to healthcare; it came out a couple of years ago and was sponsored by the Health Foundation. Our conclusion was that AI is much better thought of as a way to automate away much of the tedious admin work that plagues frontline workers in the NHS today, particularly in primary healthcare. We would like to see much more thought about how AI can help doctors to process letters and do the management of data. The famous fact that the NHS was the largest purchaser of fax machines in the world until very recently points to the need for much greater investment in AI for the routine processing of information, which we think could deliver great value.

Chair: Your evidence is attracting lots of supplementary questions around the table, but I will stick with the order that we planned. We will go to Tracey and then Aaron, and I will bring in colleagues who want to ask supplementaries at the end.

Q23 **Tracey Crouch:** Is it right to say that we don't know if or when transformative AI will be developed?



Michael Cohen: I think it is right to say that we do not know when.

Q24 **Chair:** Just before you answer that, Mr Cohen, perhaps you can define “transformative AI”. We have had a definition of AI.

Michael Cohen: Different people might use the term a bit differently. I can try to answer the question of if and when superhuman AI is likely to come, since I am a bit more comfortable with the definition of that.

Chair: Perhaps we can bring Professor Osborne in on the definition of transformative AI.

Professor Osborne: As Mr Cohen correctly said, “transformative” can mean many different things. You could very reasonably claim that we already have transformative AI in the form of ChatGPT, which is an advance perhaps comparable to that of search engines, and maybe even the internet, if you are feeling particularly enthusiastic. Certainly, the technology that we have already has enormous potential for transformation across the economy and in society more broadly.

But when we think further into the future, considering Michael’s speciality of superintelligence, there is, as he said, a lot of disagreement, but—at least to me—things seem to be really rapidly developing right now. We have this quite worrying development of arms races emerging. For instance, very recently, Google has called back in Larry Page and Sergey Brin and has said publicly that it is willing to recalibrate the level of risk it assumes in any release of AI tools due to competitive pressure from OpenAI. What they are saying is that the big tech firms see AI as something that is very, very valuable, and they are willing to throw away some of the safeguards that they have historically assumed and to take a much more “move fast and break things” perspective on AI, which brings with it enormous risks.

Q25 **Tracey Crouch:** Mr Cohen, what is superhuman AI?

Michael Cohen: Superhuman would be that, for pretty much any task we can do, it could do it better. Intelligence can be a bit difficult to think about, in terms of what that means, but the ability to complete tasks is a bit more concrete. If, across a broad range of tasks, AI is better able to accomplish them than us, I would call it superhuman.

Q26 **Tracey Crouch:** So you think that that is coming, but the uncertainty would be around when.

Michael Cohen: Yes. Certainly on our current track, it seems like there will be continued investment in AI until it does come. And no, I have no idea when it will come.

Q27 **Tracey Crouch:** Are there any implications in that uncertainty?

Michael Cohen: I think the implication of the uncertainty is that conservatism demands that we think about the worst possibilities, so I think we should prepare for it coming on the sooner end, because if it does



and we have not prepared, we are out of luck. We might get lucky and have a bunch of time, but we might not.

- Q28 **Tracey Crouch:** Could you expand on some of the risks that you think are posed by AI systems to their own users or to those who encounter them in everyday life? How do you think those could be mitigated? Professor Osborne just mentioned some risks, but could you expand a little on others?

Michael Cohen: With superhuman AI, there is a particular risk that is of a different sort of class, which is that it could kill everyone. There is an analogy that is actually pretty good. If you imagine training a dog with treats, it will learn to pick actions that will lead to it getting treats, and we can do similar things with AI. But if the dog finds the treat cupboard, it can get the treats itself without doing what we wanted it to do. If you imagine going into the woods to train a bear with a bag of treats, by selectively withholding and administering treats, depending on whether it is doing what you would like it to do, the bear would probably take the treats by force.

The way we do long-term training today with AI is broadly pretty similar to the way we train animals. If you imagine doing that, the paradigm where the AI is incapable of taking over the process will look completely different from the paradigm where it is. If the AI is able to gain power in the world and intervene in its feedback, that would actually be what its algorithm tells it to do. The output of the algorithm would look totally different from the setting where in order to get the treat, it has to do what we like.

If you have something much smarter than us monomaniacally trying to get this positive feedback, however we have encoded it, and it has taken over the world to secure that, it would direct as much energy as it could towards securing its hold on that, and that would leave us without any energy for ourselves.

- Q29 **Tracey Crouch:** My primary question was how it can be mitigated, but you have implied that it can't.

Michael Cohen: No, it can.

Tracey Crouch: Thank goodness.

Michael Cohen: Well, what I have described does not apply to all forms of AI. For instance, I was talking earlier about the economic benefits of human imitation of AI. If you are training a human to imitate AI, it would not take over the world any more than the human it is imitating would, so that is a different algorithm that gets encompassed under the very broad term "AI".

AI can cover prediction and planning mainly, and for things that are only doing prediction, this is not an outcome that I think we should expect. It is distinctly possible to develop regulation that prevents the sorts of dangerous AI that I am talking about, while leaving open an enormous set



of economically valuable forms of AI. But I think we would need to regulate away certain algorithms.

- Q30 **Tracey Crouch:** In his opening questions, the Chair asked about the lack of consensus in terms of definition and the governance challenges that we face. If there is no consensus on the definition, how can we design a regulatory system with all these different aspects to it?

Michael Cohen: I think AI is the wrong concept. There is no consensus on the definition of that, but the risks are not from AI broadly; they are only from a specific subclass of algorithms. We can have regulation that is targeted at long-term-planning artificial agents using a learned model of the world, and we can use only terms that are much easier to get a handle on. By steering away from the terms that no one can agree on exactly what they cover and focusing on the terms that are much clearer, and which actually hew much closer toward separating the good from the bad, I think we can develop national and international standards on this.

- Q31 **Tracey Crouch:** Forgive my ignorance, but I have a question about the use of black box AI systems. Is that the same as superhuman, or is that something different?

Michael Cohen: It is in theory different, although I think a lot of machine learning practitioners believe that the most capable models will be black box models—that is, you will not be able to look into them and inspect all of the things they believe.

- Q32 **Tracey Crouch:** So how can they be regulated?

Michael Cohen: Regulation that tries to inspect the models and see what they are thinking is less likely to succeed, but I think you can come up with successful regulation based solely on the design of the algorithms that are producing the black box models, rather than the inscrutable circuits that are being generated.

- Q33 **Tracey Crouch:** Could that regulator be AI?

Michael Cohen: I suppose it could be assisted by AI, but I think humans could manage.

- Q34 **Katherine Fletcher:** You are talking about setting the rules of the game for the computer program. That is what you mean, don't you?

Michael Cohen: Yes.

- Q35 **Stephen Metcalfe:** On this idea of black box and explainability, I agree that it is very difficult to build explainability into a black box, because it is a mystery. What about repeatability or predictability, so that you have some way of deciding whether what you want from the AI is what you are getting from the AI? If you cannot build in explainability, surely there must be some way of controlling it, bearing in mind that, as I said earlier, nothing is artificial about AI and nothing is intelligent about it. Back to your "Terminator" plot, surely we can just pull the plug?



Michael Cohen: Today, absolutely—we wouldn't even need to. If you have something that is much smarter than us across every domain, it would presumably avoid sending any red flags while we still could pull the plug. If I were an AI that was trying to do some devious plot, the first thing I might do is access the internet and get my code copied on some other machine that no one knows anything about. At that point, it is much harder to pull the plug.

Q36 **Stephen Metcalfe:** Let's not go down that rabbit hole too much, because I think this is an unrealistic scenario. Let's talk about the actual black box, and repeatability as opposed to explainability.

Michael Cohen: Repeatability seems like a good thing to look for. If you are wondering whether something is going to do something truly catastrophic, running it through a bunch of tests is not going to help much. So repeatability could work in some settings, but not the ones that I am concerned about and you are sceptical about.

Q37 **Chair:** Professor Osborne, you have heard the exchanges on black box AI. Perhaps you could just give us your perspective on that, and then I will turn to Aaron Bell, who has some further questions.

Professor Osborne: First, I want to endorse everything that Mr Cohen has said. I would like to introduce a framework that might help us think about some of these issues of governance, which is distinguishing the normative questions of AI from the technical questions of AI. I think good governance should consider both.

Q38 **Chair:** Sorry—normative and technical, did you say? I missed your second one.

Professor Osborne: Normative and technical—exactly right. Thinking normatively, there are some applications of AI that should just be disallowed. For instance, I do not think that facial recognition should be used by the police, to give one example. I think that should just be outlawed. We also want to have some—

Q39 **Graham Stringer:** Professor, can you expand on that? Can you say why?

Chair: Can you explain why you think that clearly should be banned?

Professor Osborne: Well, perhaps I have gone too far, but it is at least something that we should consider banning. As we have described already, these questions of bias are so difficult to erase from AI that we have to be very careful in any deployment of facial recognition in policing operations, to ensure that it was not problematically biased. If not that example, there are others in which, I am sure we would all agree, AI should not be used.

Q40 **Chair:** I am sure that during our inquiry, we will go into some detail on lots of these. For an opening session, you are giving us so much food for thought, but continue with your distinction between the two types: the normative and the technical.



Professor Osborne: Technical goals for AI consider the best practices of how AI should be used. Here we come to that question of whether an AI should be a black box, to what degree an AI should be able to explain itself and to what degree an AI should be robust to various attacks. Should its performance be subject to certain guarantees? Those two questions are slightly distinct. One is more about what we would want to agree should or should not be a rightful target for AI, and the second question is more about, in the inner details of AI, how it is best to achieve the goals that we have set.

Chair: Thank you.

Q41 **Aaron Bell:** Thank you, both. We have got to the slightly bleaker part of your written evidence, where you use words such as “catastrophic” and “existential risk”. To follow up on what Steve said, how realistic do you actually think this is, not in the next 10 years, but at some point in the future? It seems to me to have parallels with the all the concerns about the particle accelerator at CERN inadvertently ending the universe. Is this something that you could realistically see being a problem, say, in the next 30 years, Mr Cohen?

Michael Cohen: Yes. You look at it at first glance, and you can see some similarity between concerns about the CERN particle accelerator and this. And then, what academics do is look closer. They look closer at CERN and say, “No—maybe a reasonable idea at first glance, but not actually a concern.” They look closer at this and say, “It still seems concerning,” and they look even closer and say, “Yes, actually, the output of these algorithms, if they were doing planning much more effectively, would in fact behave in this way.” The only difficult part to get your head around is the fact that artificial systems really could be as good at outfoxing us geopolitically as they are in, say, very simple environments such as games; Professor Osborne was explaining just how much simpler they are. If your life depended on beating an AI at chess, you would not be happy about that.

Q42 **Aaron Bell:** What has impressed us recently is ChatGPT and the image-generation stuff. Is that basically an extension of that? Those areas are obviously creative arts, but they have game-like rules to them. Obviously, they are much more complicated than chess, Go or anything like that. Is that what has happened in the recent development—that they have found the rules for writing poems, speeches and things like that, and also have the training data?

Michael Cohen: Yes. In the last few years, I think you have seen AI becoming much more generalist in some ways. It is able to understand the rules of much more complex domains, and I expect that progress to continue.

Q43 **Aaron Bell:** Is there a reason why there would be a limit on that in any way, in terms of the complexity of the natural world, or even going down to uncertainty principles and all the rest, such that AI could not grasp those things? You seem to be quite fatalistic about the possibility of—



Michael Cohen: Okay, it is not going to know the velocity and position of a particle at sufficient precision—you mentioned the uncertainty principle—but beyond those physical limitations, no, there is no limit. Astonishingly to me, while birds were flying around them, people denied the possibility of heavier-than-air flight. We are examples of intelligent agents who can understand the world. There certainly isn't any reason to think that AI couldn't get to our level, and there is also no reason to think that we are the pinnacle of intelligence.

Q44 **Aaron Bell:** You obviously think it is a realistic prospect that an AI could emerge that seeks to gain control over our resources, our lives and so on. Overall, it looks like a fairly risky endeavour. Do you think it is realistic to try to regulate that away? That would surely take global regulation, and even then, there is no need for anyone to necessarily stay within those rules.

Michael Cohen: If we can develop a shared understanding of the risks here, the game theory isn't that complicated. Imagine that there was a button on Mars labelled "geopolitical dominance", but actually, if you pressed it, it killed everyone. If everyone understands that, there is no space race for it. If we as an international community can get on the same page as many of the leading academics here, I think we can craft regulation that targets the dangerous designs of AI while leaving extraordinary economic value on the table through safer algorithms.

Q45 **Aaron Bell:** Can I bring you in, Professor Osborne? First, how realistic do you think the bleak vision is? Are there other realistic scenarios predicting the future evolution of AI, or should we be making preparations for the bleak scenario, whenever that might arise?

Professor Osborne: I think the bleak scenario is realistic because AI is attempting to bottle what makes humans special—what has led to humans completely changing the face of the earth. If we are able to capture that in a technology, of course it will pose just as much risk to us as we have posed to other species, such as the dodo.

Of course, timelines are difficult, and there are many challenges that we might encounter either before or even alongside that. First, AI is a general-purpose technology. We have seen the world changed by the automobile and by the internet. Just thinking about language models alone, you might say that AI is already at a similar sort of scale in terms of being able to impact on a very wide variety of human endeavours. When the world changes a lot, there are risks that are posed and there are winners and losers. We have to prepare ourselves for those kinds of rapid changes.

You asked about the prospect of the international regulation of AI. There is some reason for hope, in that we have been pretty good at regulating the use of nuclear weapons—at least for several decades—where there is a similar sort of strategic advantage, if someone was able to use them successfully. If, as Mr Cohen said, we are all able to gain an understanding



of advanced AI as being of comparable danger to nuclear weapons, perhaps we could arrive at similar frameworks for governing it.

- Q46 **Aaron Bell:** If you could highlight one factor—positive or negative—that is particularly key to the future development and impact of AI, what would that be? Would it be the regulation?

Professor Osborne: If that question is for me, yes, regulation is certainly key. One thing we absolutely need to try to prevent are arms races. Unfortunately, as we speak, I think we are in a massive AI arms race—both geopolitically, with the US versus China, and among the tech firms. As I said earlier, there seems to be this willingness to throw safety and caution out the window and just race as fast as possible to the most performant and advanced AI. I think we should absolutely rule those dynamics out as soon as possible, in that we really need to adopt the precautionary principle and try to play for as much time as we can.

- Q47 **Chair:** On that, “arms race” is a metaphor. Could you unpack that a bit? What do you literally mean by what is going on that you characterise as an arms race?

Professor Osborne: It is literally an arms race in that AI is a military technology. AI can be used to control drones much more effectively and do satellite recognition. It serves various military purposes. The geopolitical actor that masters a particular AI technology first may have enormous strategic advantages, so, in that sense, it is quite literally an arms race. When it comes to civilian applications, there are similar advantages to being the first to develop a really sophisticated AI that might eliminate the competition in some way. If the technology that is developed does not stop at eliminating the competition, but perhaps eliminates all human life, we’d very worried.

- Q48 **Aaron Bell:** Just to finish, you said earlier that prediction is a mug’s game. I will not ask you to bet as such, but where do you see AI in five, 10 and 20 years—in each of those timeframes? How realistic is it that it will be positive for the next five years and then the risks will come later? Could you give us a few wild stabs in the dark perhaps?

Chair: Or wise steers to the Committee.

Professor Osborne: One thing I am certain about is that there is a lot of change coming. It is more difficult to predict the direction in which the change will come. We tried to do exactly that in our reports on the future of employment, and broadly we expect AI to be doing more routine work and less creative and social work.

As a broad framework, you can expect tasks that involve routine, repetitive labour and revolve on low-level decision making to be automated very quickly. For tasks that involve a deep understanding of human beings, such as the ones that are involved in all of your jobs—leadership, mentoring, negotiation or persuasion—AI is unlikely to be a competitor to humans for at least some time to come. Timelines are



HOUSE OF COMMONS

difficult, but I am confident in making that assessment for at least the next five years, say. Beyond that, my predictions get much murkier.

- Q49 **Aaron Bell:** Self-driving cars are an example of AI. I am struck that we seem to be a long way behind where some people thought we would be with those. Is that an example of where people have underestimated the complexity of the real world?

Professor Osborne: I think that is right. Self-driving vehicles have been predicted by the likes of Elon Musk to be on our streets within one year for about seven years. I think it is right that they haven't achieved what we had expected. There are several reasons for that.

Of course, there are technical challenges with autonomous vehicles concerning corner cases. What do you do when a cat jumps on to the road in front of your vehicle when it is also raining and perhaps there is a car coming slightly in the same direction as you? They are difficult challenges for an AI to manage while remaining safe. The technical challenges are coupled with other challenges. Some of those are regulatory. We would need to come up with a regulatory framework to make consumers willing to invest in these technologies.

There is also a need for new insurance products to give assurance to consumers before they buy. Also, there are many legal questions about the use of driverless vehicles that have not yet been fully solved. Those are some of the things that have been holding up driverless vehicles. In other areas, we have seen AI progress much faster than we had expected. We have talked a lot about language models and text to image already. Both those examples should give us some pause about the difficulties of prediction. Even with all the expertise that we have developed in understanding AI, even in very recent history, we have underestimated the progress made in some areas but overestimated it in others.

- Q50 **Aaron Bell:** Thank you. Finally, the same question to you, Mr Cohen. Where do you see AI in five, 10 and 20 years?

Michael Cohen: I have a similarly unsatisfying answer, if you can forgive me, Mike, for saying that. It is hard to predict. One historical fact is that Rutherford, a prominent physicist, famously said that nuclear energy was never going to happen, or it would happen years and years in the future. The amount of time that it took between that claim and Szilard's invention of the nuclear chain reaction was less than 24 hours. I am quite confident that nothing I am talking about is going to happen in the next week or month, let alone the next day. I think we have several years at least.

- Q51 **Aaron Bell:** But when we get there, the moment might sneak up on us.

Michael Cohen: It might look quite a lot like it does today months before. Technological progress often comes in bursts. There is a chance that we will see something like this in five years, but it is a small one. There is a chance that we won't see it for another 60 years.

Chair: Thank you. Katherine had a couple of brief supplementaries.



Q52 **Katherine Fletcher:** We tend to talk about AI as individual boxes, in a similar way to how we talk about human beings in individual boxes. Human beings are actually networks. The way our networks regulate themselves is competition for resources, which ends up as the prisoner's dilemma or game theory. Are there any AIs at the moment that can play the prisoner's dilemma correctly, to create a reciprocal, win-win, retaliate approach?

More importantly, should we be looking, when we choose to regulate concerning AI, to ensure that the resources that the AI needs to exist are valued within the game? Anything that is smarter than us, which does not perceive itself at risk for competition of resources, is going to cause a problem, because it will press the big red button on Mars. The resources are electricity—the energy needed to input—as well as the maintenance of external components, so can't we just put those in?

Chair: Can we have a brief response to that, and then we need to move on to our next session.

Michael Cohen: I will just say that if you make the setting simple enough, AI can solve the prisoner's dilemma. Could it figure it out in the real world? There are settings complicated enough where I am sure that AI today could not do it. But I would expect that as AI advances it will do the game theory-optimum things.

Q53 **Katherine Fletcher:** What about competition for resources to regulate the AIs?

Michael Cohen: I am afraid I did not quite follow the question.

Katherine Fletcher: Basically, AI now does not worry about where its energy is coming from to run, and it does not worry about the fact that the pipe is rusting—in a similar way to how we need food and air to breathe.

Michael Cohen: Right, okay. If it got sufficiently advanced to realise that any electrical failure would pose a problem to its future ability to achieve its goals, then it would start to be concerned about that.

Q54 **Katherine Fletcher:** But wouldn't you regulate it to tell it that that is its problem in the first place, so that it knows it is beholden to a higher master?

Michael Cohen: That sort of regulation I would be less confident we would be able to craft.

Chair: We will come on to that in subsequent sessions. Thank you. We have run a little beyond time, but that is a reflection of the degree of interest in your evidence. Professor Michael Osbourne, and Mr Michael Cohen, thank you for your written evidence and for answering questions so fulsomely this morning.



Witnesses: Katherine Holden and Dr Manish Patel.

Q55 **Chair:** We come to our second panel of witnesses. I welcome Dr Manish Patel, who is the chief executive and technical architect of Jiva.ai, which is an AI firm specialising in healthcare applications. Joining us virtually we have Katherine Holden, who is the head of data analytics, AI and digital identity at techUK. Thank you for joining us.

I will start with a question to Dr Patel. Thank you for your patience in sitting in the room for the first panel. Could you briefly describe how the firm you set up, Jiva.ai, uses artificial intelligence?

Dr Patel: Thank you for having me. The general idea is that we are trying to make AI easy to access for those who are not technically able to create it. There is a problem, I think, especially in the UK, where there are not enough data scientists, AI engineers and people who have actually worked with this type of technology to do good things, such as create diagnostics and automations for healthcare. We are trying to put together technology that makes that easy to use for non-technical people.

In the same way that, 20 years ago, you might hire a HTML developer to do a website, but now you use WordPress or Wix, we are trying to put a technology together where someone can come and just describe what they want to create.

Q56 **Chair:** As I said, you obviously operate in the healthcare and medical spaces. Can you describe what you see as the benefits of that—which I assume are the same as what your firm offers?

Dr Patel: Sure. The last session touched on some of the more difficult areas of bias and regulation. The very obvious application is diagnostics. There are a lot of companies out there trying to create cancer diagnostics. For example, they take an image of your skin or take a snapshot of an MRI or a CT scan and try to determine whether you have a particular type of cancer—it's that kind of thing.

Then there are types of applications that are less directly applicable to patients. An example is automating administrative tasks in hospitals. One of the customers that we are working with at the moment is simply trying to clean up the procurement data they get from a hospital, so that they can work more efficiently and have the hospital get their budget quicker. So there are various types of applications, but ultimately the benefit is a more efficient healthcare service and more efficient service to patients.

Q57 **Chair:** In terms of the work that you do, what is the balance between those two things? One is administrative improvements in efficiency, and the other is greater diagnostic accuracy—I suppose one might characterise it as that.

Dr Patel: That's a great question. Up to, I would say, 2020-21 and during the pandemic there was a lot of focus on diagnostics. People maybe got a little bit carried away with what AI and machine learning could do in that particular area. I think that, more recently, people have come to realise that it is actually quite difficult to prove that your diagnostic works in a



clinical setting. The MHRA has done a really great job in putting together the framework to ensure that safe release of this technology happens. Therefore, the more commercially minded people out there have decided to go for the lower-hanging fruit, in terms of automation of more clerical work. So I would say there is a little bit of a move away from diagnostics at the moment, but in proportion rather than in number.

Q58 Chair: I see. Your own PhD, I am told, was in mathematical and computational modelling of cancer, so you were in the diagnostics space. Is that what drew you into an interest in AI?

Dr Patel: Yes, that's where I came from. Back then, we didn't really call it AI—it was just called maths, right? Calling it AI is just a more recent thing. Yes, that is precisely where I came from, and I have always had an interest in actually applying it in healthcare.

Chair: Thank you very much indeed. I will now go to my colleagues, starting with Carol Monaghan.

Q59 Carol Monaghan: Dr Patel, you have talked about some of the benefits of AI in healthcare. What are the risks?

Dr Patel: The most direct one is incorrect regulation of a diagnostic, I would say. I will give you an example. I was speaking about a year and a half ago to a company who were working on explainability—this is the kind of thing that the guys were talking about before—and they were looking at an algo that was looking at MRI images of a breast and making a diagnosis of whether the person had breast cancer. It was getting really great results, but when they dug into it and did the explainability analysis of it, they found that actually it was just looking at one little corner of the image, where the result was actually burned into the scan.

Carol Monaghan: And the result was what?

Dr Patel: It was literally the letters of—

Carol Monaghan: Oh.

Dr Patel: So it was actually on the scan, yes. When they did this analysis, they saw that actually it was not looking at the hotspot of where the tumour is at all.

So I think the risk is that you can get carried away with great results. You can say, "Oh wow, I've got 99% sensitivity and specificity"—really great accuracy—from this particular diagnostic, but you don't know what it is actually looking at.

Q60 Carol Monaghan: To go back to the example you have just given, that wouldn't be 99% accuracy, because it's picking up something wrong entirely. That wouldn't be the great results that you're describing.

Dr Patel: No. Accuracy, I guess, is relative to what you think ground truth is, and in the medical field, ground truth is very difficult to get at, even when you have the data. One of the MPs was talking about the bias within

the data. Within healthcare data, you have the curation of that and the decision making that goes into what you think is a diagnosis, but even the techniques that you use to do the incumbent diagnosis are flawed, and therefore you don't whether you have actually got ground truth.

Q61 **Carol Monaghan:** You have talked about AI helping to interpret CT scans, for example, to look at cancer diagnosis. Who is overseeing that that's right?

Dr Patel: There will be various organisations that you would have to go through to get something like that—as an example—into the NHS.

Q62 **Carol Monaghan:** So it is not being used just now.

Dr Patel: No. Let us be clear. Universities and university hospitals can do R&D as long as the patients have consented to their data being treated in that way to do those kinds of things as part of a clinical trial, for example. If I or one of my paying customers on Jiva were trying to create a diagnostic of disease X, they would have to go to the MHRA, get it regulated and have a QMS system. It would be regulated as a class 2A, 2B or 3 medical diagnostic. That would involve their going through a number of clinical trials to build the evidence to get to a point where the MHRA or the FDA signs off and says, "Okay—"

Q63 **Carol Monaghan:** At that point they are confident that that can be used.

Dr Patel: Yes, up to the point at which the governmental organisation is confident, but there is a very high barrier to entry for these companies to get into the NHS or any healthcare system globally, because the doctors on the other side will say, "Let's have a look at your training dataset. What do you have in your training dataset? If you have 90% men and 10% women, that is not very representative of the population. That is not something that we can work with." So there are multiple levels between having an idea, doing it and creating an AI to its being usable in the clinical context for patients.

Q64 **Carol Monaghan:** Okay, so in that example we are a bit off. In 10 years' time, what can we expect to see in terms of AI in healthcare?

Dr Patel: Someone said it quite succinctly. I don't think we are looking at a situation where AI is replacing humans. It is not replacing doctors or radiologists. It is definitely an augmentation tool. I don't see that changing in the next 10 years. I think the medical community has to be confident that this technology works for them, and that takes time and it takes evidence.

Q65 **Carol Monaghan:** Going back to the example you have given, we have used an AI tool and looked at a CT scan, or the AI tool has examined a CT scan. It has picked out potential areas of interest or concern. I assume that at that point a healthcare professional comes in and checks that. Why do we need AI?

Dr Patel: Because it makes it faster. I will give you an example. We work in the prostate cancer space. There is a particular issue in prostate cancer

where the vast majority of diagnoses are indeterminate at the MRI stage. We give it a 3 score, for example, with an average score to say it could be something but could not be something. We are not sure, so let us have a biopsy.

When you put that patient through a biopsy and you have a false positive, meaning they do not have cancer but they have gone through a biopsy, they are going to go through the mental stress for six to eight weeks of waiting for that biopsy and then waiting for the result. They will have the side effects of that biopsy and things like erectile dysfunction. That costs the NHS a lot of money because those men are generally 65-plus and susceptible to infection. They will therefore come back with all sorts of issues—sepsis and that kind of thing. What we are trying to do with an AI is to say: in the clinical pathway where you are a man of a certain age, you have an MRI scan. Well, have the AI take a second look. In your radiology list you have 150 guys on that list every week that—

Q66 **Carol Monaghan:** You have gone through all these scans. You have not picked anything up, but you throw it into the AI.

Dr Patel: Throw it into the AI and it can stratify. We can say, “These are definitely like a 5. We know this looks really dodgy, so prioritise these people.” It makes their MTT session in the pathway a little bit more efficient.

Q67 **Carol Monaghan:** Is there anything else on the horizon that we should be aware of that is exciting or concerning?

Dr Patel: Yes. I think there are certain applications in the primary care area that are particularly interesting for us in the UK where there is an incredible amount of strain on GPs. There are certain technologies now coming out that will help them be more efficient in putting referrals through to the right areas, and help them understand where the pressure points are in the NHS between the interface of primary care and secondary care. There is also a similar situation between secondary care and social. The efficiency gains that we can make should be really quick and easy—well, not that easy—and there is only a benefit to doing that. I don’t see any downsides.

Q68 **Carol Monaghan:** Finally, are there any things that you are concerned about, in terms of risks of AI being used in healthcare, either now or in the future?

Dr Patel: Aside from the regulatory one that I mentioned?

Carol Monaghan: Maybe you could expand on the regulatory one. Is it a lack of regulation that you are concerned about?

Dr Patel: I think there is a lack of understanding. The UK has been really great with life sciences in the last 40 years. We have a really great base. There are more life sciences companies in the UK than in the whole of Europe combined, and all of them want to use AI to some degree. That could be in drug discovery or diagnostics—it could be anywhere. When they go through the regulatory framework, the people on the other side of



HOUSE OF COMMONS

the fence don't necessarily have the skills to understand what is going on, and therefore to understand whether it is something that needs to be regulated in the first instance, and if it does need to be regulated, how. I guess that this Committee exists to establish that.

The risk is that you get an AI—something that is either doing a diagnostic or automation—that inadvertently puts patients at risk. That is the big one. If an AI puts one patient at risk and that hits the *Daily Mail*, that has way more weight than a doctor making a mistake.

Carol Monaghan: Thank you.

Chair: Thank you very much indeed, Carol. We next go to Katherine Fletcher.

Q69 **Katherine Fletcher:** Thank you very much for coming. Apologies for getting over-excited at the end of the last session. Despite some of the apocalyptic things that we need to regulate properly, we need to embrace the opportunity that AI provides the UK. Katherine—I am delighted to welcome another one with your name spelled correctly—perhaps I can turn to you first. I want to understand what you are excited about with AI in the UK economy, and what the main benefits are of increasing its presence in this more point-and-click format, which everybody is quite comfortable with.

Katherine Holden: Thank you very much for the opportunity to join you today for this hearing, and for pivoting online at the last minute. It was very much appreciated.

From our perspective at techUK, as the trade association, we obviously see AI as being hugely important to the UK economy and our wider society. To give you some stats, we expect that AI could deliver a 10% increase in UK GDP by the end of 2030. As Dr Patel and others have mentioned already as part of this evidence session, we have seen some notable commercial AI successes in this space already, such as in health, pharma and financial services, to name a few. We are still very early in our AI adoption journey here in the UK. Although 30% of businesses in the finance and legal sectors have adopted some form of AI, only about 12% have done so in the retail and hospitality sectors. I definitely think there is an appetite and an opportunity over the next few years to really look at how this technology can be adopted more broadly across lots of different sectors.

Q70 **Katherine Fletcher:** Could you perhaps give some examples of that? Thanks to Dr Patel, we have some ideas about what health and pharma can do. What is going on in the financial and legal spaces with AI? What opportunities are there? That 30% figure is quite striking.

Katherine Holden: Yes, absolutely. Particularly in the financial services sector, one of the opportunities we are seeing—this is still fairly nascent—is to use AI to identify fraud in many different forms. Obviously, that is incredibly important, because we are seeing millions and millions of pounds every year lost to various instances of fraud. AI has the ability,



HOUSE OF COMMONS

particularly in its application within the banking sector, to identify examples of fraud—perhaps the imitation of individuals—very quickly, and obviously in real time. That is one opportunity we are seeing.

At the moment, a lot of the opportunities associated with AI can be seen as fairly mundane and in areas where you may not even notice they are part of your everyday life. We mentioned optimising our journeys, making sure we have relevant search results and so on. For most of our members—we have 1,000 members at techUK—the benefits of AI truly come back to improving business productivity and streamlining business operation.

To give you one example that was particularly important during the pandemic, one of our members, Ocado, used a smart algorithm that was able to identify that temporarily switching off bottled water basically saved a huge amount of space in their delivery vans, and they were able to serve an additional 6,000 customers a week. During the pandemic and the lockdown, that meant that some of the most vulnerable members of society were able to get access to food and resources, and it helped to prevent the spread of the virus in the supermarket. That is one example among many, across all different sectors and markets.

- Q71 **Katherine Fletcher:** That is the AI spotting the calculation of a product's relative profitability based on the volume it takes up in the van and recommending certain lines be pulled from the van to create more space.

Katherine Holden: Yes, absolutely.

- Q72 **Katherine Fletcher:** That is exciting. Where does the UK stand on this in a globally competitive marketplace? I will not use the words "arms race", because that is different, but the opportunity to do business more efficiently is something we all get a little excited about.

Katherine Holden: The UK stands in good stead when it comes to AI. We have an absolutely thriving AI economy here in the UK, from the big tech firms to a number of agile, exciting smaller SMEs, start-ups and scale-ups. As a sector, we are thriving. Only China, the US and possibly India are ahead of us in terms of investment. The area where we need to really focus our efforts now—and I am sure we will come on to this—is the importance of AI governance and AI regulation. If we get that right, it is a huge opportunity for us to continue to be an ever-successful global player in this area.

- Q73 **Katherine Fletcher:** Thank you for that, Katherine. Dr Patel, what made you come to the UK for the AI ecosystem and to take the opportunities here?

Dr Patel: Well, I am here, so it is easier.

Katherine Fletcher: I did not mean you personally.

Dr Patel: There are a couple of things. There is a lot of experience in the team around healthcare, which is why we targeted healthcare to begin with. The NHS is ripe for uptake. It is also a great fertile ground for doing



R&D, because there are very close affiliations between hospitals and universities. That is fairly unique to the UK. I mentioned that the UK has a very vibrant life sciences sector, which places it in a great position to lead the charge on AI in healthcare and life sciences. If we were not to capitalise on that opportunity, we would be missing a trick.

Q74 **Katherine Fletcher:** Can you give us one example of why we are so well placed? We hear a lot about how we are world-leading in life sciences. I always try to granularise what we talk about in the Committee.

Dr Patel: Sure. I assume you have heard of Genomics England. At the tail end of the human genome project, they were really flying—well, their predecessor was. For that to occur, there was a recognition that the human genome project was great and got loads of data, but to translate it into medicine and actual human benefit, there is a whole lot of analysis that you need to do. Therefore, there was the birth of citizen biology and bioinformatics, and again, the UK was right at the front in 2004 and 2005, doing that work. That is a quick example of UK academic institutions, hospitals and big corporates understanding where the opportunity was and jumping on it and capitalising on it.

Q75 **Katherine Fletcher:** Okay, so it is more opportunity than mechanisms, standards and regulations that is the primary attractant.

Dr Patel: In any field, the R&D bit comes first, and then everybody else follows. No one was going to regulate nuclear energy before it became a reality or we thought of it becoming a reality. It is the same with AI, especially in healthcare. We have suddenly realised you can do all of these things. Yes, there are problems; there are problems with data and application of it, and that kind of thing, but that R&D has to take place first. That R&D takes place in two different areas: academic institutions and start-ups. Those are small agile companies that have the ability to do those innovative, creative things. If that does not exist, you do not have the regulations and you do not have the benefit at the end of it.

Chair: Before I bring in Dawn, Aaron you wanted to follow up on a point.

Q76 **Aaron Bell:** Yes. Ms Holden, maybe I did not catch it right earlier, but I thought you said that AI could grow the UK economy by 10% by the end of 2030. Was that what you said?

Katherine Holden: Yes.

Q77 **Aaron Bell:** That is quite a remarkable claim. That is an extra 1% growth a year over the next eight years. What is that based on? How realistic is that, when we have struggled to get much more than 1.5% growth over quite a long period of time? It's the productivity puzzle.

Katherine Holden: This is a piece of research that was conducted on behalf of Government. I am happy to share resources and links post this session. It goes back to where we are in terms of the adoption journey in the UK.



We have seen examples of successes within some of the major markets and sectors, but there is still a huge amount of untapped potential, and some quite quick wins if we ensure we get the AI governance and regulatory system right in the UK. It is one of those areas where we can see quite a lot of low-hanging fruit and quick wins that could happen in a very short period of time.

Q78 **Aaron Bell:** Is there some growth already banked from AI that we can point to, to demonstrate that this is realistic over the next eight years? In the last few years, what proportion of GDP growth could you attribute to AI?

Katherine Holden: I would have to come back to you with some details on that. I don't have that.

Q79 **Chair:** Would you write to the Committee on both points: the origin of the 10% figure and the other impact?

Katherine Holden: I am more than happy to.

Q80 **Dawn Butler:** I have a quick question for Dr Patel. Thank you for coming in today. Can Jiva.ai out-program inherent biases that already exist? For example, body mass index is discriminatory. It was created by a mathematician who identified the average man. It did not take women or different ethnicities into consideration. Yet, it is still used medically every single day to determine outcomes for patients. Does your program in any way out-program any inherent biases or does it continue them?

Dr Patel: Not automatically, is the straightest answer I can give you. You quite rightly pointed out the inherent bias within BMI. There are lots of examples of that across the medical field—things still in use today that we know are just not completely correct. We are striving to highlight those issues in the data.

If you were to gather a primary care dataset that had the raw material data, if you like, that contributed to those higher-level calculations, such as BMI, our system might be able to point out that there is a particular bias here and that is why you get this particular result. We rely on the human on the other side to look at that and understand that there is a bias there—that there is something not quite right in that particular predictive analytic.

It is even more problematic for us to do that automatically than to point that out to the human, who can say that there is something there. I would be reluctant to go down that route, to effectively automatically codify a bias out, which is not an easy thing to do anyway.

Q81 **Dawn Butler:** Have you identified all of the bias? It is almost impossible, I know—sorry. Have you identified many of the biases that currently exist within healthcare?

Dr Patel: Our customers would. If someone comes to us and says, "We want Jiva to help us create a diagnostic for disease X," typically our first few calls are about what available data they have and if it is really



representative of what they want to do. When they go through the process and create a model, even before that model gets to a point where it is in front of patients there will be a health economics analysis and a clinical trial, where those items will get weeded out. We have seen time and again—not in Jiva’s lifetime, but in other applications—instances where people have come forward and put these diagnostics together, and in the clinical trial they have been able to weed out the missing components and the bias in that data in the first place.

Dawn Butler: Thank you.

Q82 **Stephen Metcalfe:** We have obviously heard about the dystopian future that AI will deliver, whereby the machine takes over and we are all consigned to slavery. I do not necessarily buy into that being the outcome, but to make sure that it doesn’t happen we obviously need some form of regulation. When thinking about the regulation of AI, what should Governments consider to ensure that we continue to develop AI for the general good and not leave the risks in place that could come back and bite us in five, 10, 50 or 100 years from now? Who would like to start?

Katherine Holden: I am very happy to start. More broadly, I know the UK Government is imminently producing its UK AI White Paper. I will just put a stake in the ground and say that techUK very much supports the Government’s initial proposal for what we would describe as a pro-innovation, context-specific, risk-based approach to regulating AI that very much leans on the existing expertise across our regulatory regime. There are a few really fundamental things that we need to get right as we develop this proposal. If we are to rely on the existing structures we have in place, it is integral that we ensure the regulators have sufficient capacity to govern AI effectively and the ability to identify high-risk AI applications. That is key. Coherence in the development of that approach across the regulators is essential to driving innovation.

The previous panel spoke about some of the more high-risk applications. Let’s take live facial recognition technology as an example. As we look at how we could regulate or govern that technology going forward, it is really important that we establish formalised structures to co-ordinate approaches between regulators—for example, something like an expanded version of the Digital Regulation Cooperation Forum. I think the way forward is having slightly smaller working groups, where you could bring together the likes of the Biometrics and Surveillance Camera Commissioner, the ICO, probably the Home Office, and any other regulators that are impacted by such technology. That is how we stop that dystopian future—by ensuring that we have a greater level of capacity and coherence in our existing regulatory regime.

Dr Patel: I echo what Ms Holden said and agree with it all. Let us think about there being two categories of AI; I know we touched on the definition.



First, let us just think of AI as a predictive analytical tool, where you have those things that are quite low risk—that suck in some data, make a prediction about it and do not really actuate anything other than making that prediction.

Secondly, you have the area of AI that is far deeper and working towards a more human-like intelligence. That is the area that needs a more rigorous approach through regulation—without stifling the R&D process. That is really important, because to maintain an edge as a country we cannot stop people doing certain things, although I don't see how you would stop people doing R&D here anyway. You want to regulate in a way that does not stifle innovation but, when it comes to a point where that AI is productionised—out in the wild—you can intervene and say, "Okay, this AI has to operate within this particular environment." I say that really broadly because it is very difficult to speak about specifically, but that is the separation I would make.

Q83 Stephen Metcalfe: Okay. Do you think that our regulators that exist at the moment have both the expertise and the capacity to be able to take that lighter-touch approach of not stifling R&D but intervening where necessary?

Dr Patel: My limited experience of it—I mostly see the healthcare angle—is that there is the expertise, but they are highly constrained. We just do not have enough people on it, and we just do not have enough resources on it.

Q84 Stephen Metcalfe: Okay. So we need to build expert capacity in understanding AI.

Dr Patel: Absolutely. Let's take the MHRA as an example. I do not think that means growing the MHRA to 1,000 people; I do not think that is the right way of doing it. The MHRA is very capable of doing what it does at the moment. There could be a more collaborative process, where the MHRA can rely and lean on the amazing expertise we have in the universities that we have here, to spread the work.

Stephen Metcalfe: Katherine, did you want to come in on that?

Katherine Holden: Yes, just to echo Dr Patel's points—we are in complete agreement there. The reason it works so well, looking at addressing these issues through the regulatory level, is that the regulators obviously understand the context of how AI is being deployed within their own sectors and the kinds of harms that can occur. In addition to that, they also have the best understanding of the existing rules and requirements that are in place and exist already, and therefore what may need to be built on or where future regulation may be needed.

When I talk to our members, a huge problem at the moment, particularly for our SME base, is that there is a tremendous amount of guidance, regulation and standards out there, a lot of which is very much overlapping. There are areas where there are gaps, and trying to interpret that—particularly for our SMEs, who do not necessarily have that resource



and bandwidth—is confusing and complex, and it stifles innovation. Therefore, one thing we have called for in our response to the Government consultation is the need for this mapping exercise, to say, “This is what the regulatory framework looks like. Here are where there are areas of overlap, which maybe need addressing, but, most importantly, here are the gaps and these are the areas that need to be plugged to ensure that we can build greater confidence and, importantly, trust in the use of AI across all sectors and markets.”

Q85 Stephen Metcalfe: Following on from that, do you believe that as you do this mapping exercise and you see where the gaps are and what needs plugging, there should be a single Department or regulator responsible for all AI governance, to make sure that those gaps do not emerge? Or do you think that mixed landscape is useful?

Katherine Holden: I think the mixed landscape is useful as long as things do not fall through the cracks. You may need an allocated body to help oversee that mapping exercise. Again, I go back to live facial recognition, but it is always helpful to have a particular example. Say you identified that there were gaps in that area, which you may do. It would then be useful to make sure there is a body—maybe it is, say, the Office for AI—who is able to convene the right regulators together to look at how they plug those gaps in a coherent and co-ordinated way, and perhaps you have one of those regulators in the mix who takes the lead on that particular application of that particular technology. How that could work has been practised, but you would probably have to iterate it over time and see how it works in practice.

Stephen Metcalfe: Manish, did you want to add anything to that?

Dr Patel: No.

Q86 Stephen Metcalfe: Perfect. In that case, this is my final question. We have mentioned live facial recognition a couple of times, and it tends to be framed in negative terms. How much of a conversation do we need to have with the public about some of this technology? The potential flipside of the negative is that if you were trying to find a lost or abducted child in a large crowd at a football match, where the whole crowd had been surveilled, you may be able to do that very quickly, which would be a benefit of that live facial recognition. I am not suggesting we roll it out; I am trying to make sure that we frame the discussion within the pros and cons of certain technology.

Katherine Holden: Absolutely. What is important here is defining exactly what we are talking about. Often people use the umbrella term of biometrics, for example, but that can be a term for many different technologies and here we need to be very specific. We are talking about live facial recognition in the context of policing—that is enough ethical issues and areas to tackle in one go.

A lot of work has to be done in that space to build public trust and confidence, and wider conversation needs to happen among regulators about where the gaps need to be plugged. In Europe, we mostly see a full



ban on the technology, which I am not sure necessarily is the way that the UK wants to go because, as you rightly said, that could stifle innovation—there are also applications such as finding missing children or in the healthcare space to identify mental health illnesses. We need to make sure that we don't throw the baby out with the bathwater.

We are looking at the opportunities that exist and making sure that we regulate accordingly. Of course, as part of that, it is essential that we bring along the public as part of the conversation. We must make sure that we are communicating clearly where the technology can and cannot be used, with a wider public engagement exercise and education piece on why the technology is important, if regulated in the right ways and applied in the right sectors.

Q87 **Rebecca Long Bailey:** The Government are working on the AI White Paper, but would you highlight any international approaches to the regulation of AI as particularly well designed?

Katherine Holden: I have a couple of things to mention in that area. In the UK, it is worth signposting the AI standards hub, which is a collaborative piece between Turing and a few other bodies. That is an essential piece of work, because essentially it helps organisations to get a better understanding of the current and overlapping standards and guidance in the international space.

That is also important because we have left the European Union and maybe we do not have as much access to some of the negotiations in that space at the EU level. The hub highlights upcoming opportunities to influence and input into those standards. So, I would start with highlighting the AI standards hub as something that research organisations and industry across the UK should continue to feed into to ensure that we are influencing at that international level.

As for other examples, the one that strikes me and that a number of our members have mentioned is the piece of work that the Singaporean Government are doing, called AI Verify, an AI governance testing framework and toolkit—I am sure there is a better way of phrasing it, but that is essentially what it is. What is interesting and innovative about that approach is that industry has an ability, through a series of technical tests and process checks, to demonstrate their deployment of responsible AI directly to Government. It is still very much in the early stages and, at this moment, a minimum viable product, but I would recommend that the UK Government keep tabs on it and continue to look at how that technology and approach evolve and grow.

Dr Patel: I agree with all that Ms Holden said. I cannot point to any specific international regulation that is anywhere near mature enough for us to talk about in any way substantially. Everyone is still at that foundation-building phase. We are still trying to get our heads around what it is and what the risks are, and building the regulation around that.



I cannot speak more about that, but I will make a related point: there is very good regulation on data. We have inherited GDPR from the EU. I am not sure where all that is going in the next couple of years, but we have pretty good data-protection regulation. That is very closely related to the innovation piece for AI. In healthcare, we are always talking about whether we are allowed to use patient data and what we need to do to get consent from patients. For example, during covid, certain ethnic groups were more susceptible to being infected than others, and it might have been better to be able to access that data and do the analytics. There is a related discussion around building on our current data regulation as well as the AI regulation.

- Q88 Rebecca Long Bailey:** Are there any particular countries across the world that we should not be trying to emulate because their regulatory approach is poorly designed?

Dr Patel: I do not know if I will get into trouble for saying it. There are certain south-east Asian countries where you can grab data and there is not much regulation at all around it, and that is mainly because there is not fundamentally a respect for people.

- Q89 Rebecca Long Bailey:** Katherine, may I ask you the same question? Are there particular parts of the world whose regulatory approach we should not be trying to emulate?

Katherine Holden: I agree with what Dr Patel said. Without going into the specifics, one jurisdiction that it is worth being aware of—I am not saying it is necessarily a bad approach, but it is a different one—is the EU. They are looking at taking a very different approach to how they tackle AI regulation, with a very specific piece of AI legislation: the EU AI Act. They are looking at regulating a definition of AI, while the UK is looking more at potentially regulating the outcomes of AI. That is obviously a very centralised approach, with some very hard-line regulation. From my perspective, it does not allow the opportunity for much flexibility. It is not particularly future-proofed, because they have a static list of high-risk applications that will be under particular regulatory scrutiny. We know that a static list would need updating very, very regularly. There are many other issues about having a blanket AI piece of regulation, and I can go into more detail if necessary.

The thing to be aware of is that a lot of the companies operating here in the UK are also operating in Europe. That is a huge market for a lot of our members. As part of the EU AI Act, there will be extraterritorial reach, which will mean that any AI system that is providing outputs within the EU, regardless of where the provider or user is located, will be subject to the EU AI Act. As we are developing here in the UK our own response to regulating or governing AI, it is absolutely integral that we make sure there is a level of alignment between the work we are doing in the UK and what is happening in the EU and other jurisdictions.

- Q90 Rebecca Long Bailey:** Thank you. This is my final question. Would you prefer to see an emphasis on international regulatory co-ordination, or is



HOUSE OF COMMONS

an element of regulatory competition desirable? Katherine, you have partly answered the question already.

Katherine Holden: There is no harm in the UK carving out a very pro-innovation approach to AI that is proportionate and risk-based. It very much highlights to organisations across the world, “If you come to the UK, this is the type of environment you can expect to be met with.” That is a great thing for the UK, but we know that none of these technologies are limited to UK use only, so it is a balancing act of both.

It is integral to the success of even just the UK AI market that we have clear harmonisation with other approaches across the world for many reasons. It is important for ensuring that we do not have a fragmented global market, for ensuring interoperability and—this is really important—for promoting the responsible development of AI internationally. It is worth saying that, despite the divergence in current approach to AI regulation between the EU and the UK, we still have a very similar ethical underpinning. The way we approach issues around transparency, explainability, safety and security is very much in alignment. That may be the starting point from which we can ensure greater alignment in the future.

Dr Patel: I do not think I can better Katherine Holden’s answer. She put it really succinctly and I agree with it all.

Q91 **Dawn Butler:** Katherine, I think there is a bit of a contradiction in some of what you were saying with regards to international regulation, while criticising the EU’s approach to AI.

Katherine Holden: Okay. Would you mind explaining in what way?

Dawn Butler: You said that you do not like the way the EU is going, and that the UK can be more pro-innovation. For example, you sound quite excited about facial recognition, whereas I am quite sceptical and scared about it. I think it is fine if you have got a picture of somebody and you are trying to find them in a crowd, but that is very different from identifying somebody who the police have determined is a criminal. We have seen cases, such as that of Randall Reid, 28, who was jailed for almost a week for a robbery he was nowhere near. Is it possible to regulate out bias and discrimination from facial recognition?

Katherine Holden: It is a good question. To clarify on the EU point, apologies if that is the way I conveyed it. It is not that I am completely anti the EU’s approach, it is just a different approach, and I would say the UK’s approach is more aligned with the pro-innovation, proportionate response. The EU has a way of doing things, and it will take time to evolve before we see who has taken the better stance on how that technology should be developed.

When it comes to live facial recognition, that is exactly the type of use case that Government need to be focusing on—you are right. It was mentioned in that last session that bias is integral to most training sets. You can do lots of different things, which we can go into in more detail if



HOUSE OF COMMONS

helpful, to try and mitigate or reduce the risk of bias. But in the case of any kind of decision making, whether it is human or machine, bias is just innately part of training data.

With live facial recognition there are opportunities, but those opportunities are very narrow and context-specific. With examples like the one you mentioned, we cannot have the technology being used in that kind of way. That may be one of the ways in which we see future regulation in this space. The great way that the UK is doing it, rather than having an AI-specific piece of regulation—which would have to go through Parliament and the legal processes that come with that—is by using the existing structures across Government. That basically means we can start tomorrow. Once this White Paper is out, we can convene the regulators to spend their time on identifying what those high-risk applications may be—like live facial recognition—and making sure that we are regulating or governing accordingly. I think that it is the most efficient and effective response that we can have at this stage.

Q92 Dawn Butler: Thank you. What would you like to see in the Government's White Paper, and have you been consulted?

Katherine Holden: Yes, we have been consulted. We have met with the DCMS and other teams many times. That process has been very positive. I think they are broadly heading in the right direction.

The four key things, which I have mentioned already, are: the capacity and bandwidth of regulators and making sure they have the opportunity to focus on those high-risk applications, like live facial recognition; the need for formalised structures to ensure coherence and co-ordination across all regulators; the need to encourage and oversee the development of an effective AI assurance market, which is a key area we have not touched on; and, finally, the point I mentioned around the need for ensuring UK regulation is compatible with other approaches around the world. Those are our four key things that we feel Government should focus on.

Q93 Dawn Butler: Brilliant. Dr Patel, the same to you: have you had communications with the Government, and what one thing would you say needs to be enshrined in the White Paper?

Dr Patel: Not directly, but there has been a consultation through the BIA.

Q94 Chair: Is that the BioIndustry Association?

Dr Patel: Correct. I am far too small to be consulted by the Government for anything like that. As for the one thing I think needs to be in that White Paper, I can't select just one thing. There are actually two things. The first, with respect to healthcare, is access to data—how that is regulated and how that feeds into AI innovation. There is also the efficiency with which we can put AI technology out on the market in a way that respects patients and the outcomes for those patients, but also respects the knowledge of our great medical community. If we keep those two things separate—if we go head-first into creating diagnostics, as an example, without bringing in the great stuff we have in universities in



terms of medicine and life sciences—we will be missing a trick there. I think they need to be done in concordance, and that goes for pretty much any way you apply this particular type of technology, whether it is facial recognition, diagnostics or whatever it might be.

Dawn Butler: Thank you both very much.

Chair: Thank you, Dawn. I think Katherine had a brief question.

Q95 **Katherine Fletcher:** Thank you, Chair. Those were great questions from Dawn. I want to focus on the practicality of putting technology in with regulation. Ages ago, I was a terrible HTML coder—I loved your analogy of the fact that you don't need to be going, "Oh crap, I've missed the backslash. I've put it in the wrong place." I just want to check whether, in the practical application of AI, there are any landmines that we shouldn't stand on. In my quite archaic way, I would ask: is there any common kernel code that we need to be thinking about? What we don't want is an industry that gets really well developed, and then the regulators come in and say, "Actually, you need to do this," and you have to dig to the bottom of the coding pile to fix it.

Dr Patel: That is a really good question. Let's break down that down a little: around 93% of the entire AI community use neural networks or some flavour of neural networks—perceptron-based learning—to do the type of AI that they are doing; the rest might be using Bayesian or other types of techniques. If you focus all of your energies on regulating perceptron-based learning, which is, by nature—we talked about black box AI before—very difficult to unpick, and then we get a piece of R&D that says there is another type of algo that is not perceptron-based at all, which is much better at doing that particular type of predictive analytics, then suddenly your regulation is not applicable.

I don't think there is a kernel that you can drill down on. If you keep it really high level and say there are predictive tools where, essentially, all you are doing is pattern recognition—that is all we are really talking about, the bottom line is just pattern recognition—and then there is the type of AI that is becoming more intelligent, in terms of getting closer to human intelligence—

Katherine Fletcher: Planning, yes.

Dr Patel: Yes. If we keep it at that high level, that is as far as you need to go, but I don't think you can drill down to a particular kernel. That would be really difficult. I think that would be really painting yourself into a corner.

Q96 **Katherine Fletcher:** Okay. Very quickly, say that we wanted to say outcomes-based, which means, for example, that every piece of AI that has any planning capability within it has to understand, as one of its core programmings, that there is an existential threat to it—that is, us pulling the plug on it. Does the industry need to take account for creating those front doors or back doors, if we should need to go down that route?



HOUSE OF COMMONS

Dr Patel: Inherently, you probably will at some point.

Katherine Fletcher: Yes, I thought that might be it. We haven't got the time to explore that—I just wanted to get your sense of it. Thank you.

Chair: I am very grateful to our witnesses on both panels, latterly Katherine Holden and Dr Manish Patel. We have kicked off in fine style what is going to be a fascinating inquiry. We will be delving into all of the issues raised, and some more besides, over the next few weeks. That concludes this meeting of the Committee.