HOUSE OF COMMONS

Science and Technology Committee

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

Wednesday 22 February 2023

Ordered by the House of Commons to be published on 22 February 2023.

Watch the meeting

Members present: Aaron Bell; Dawn Butler; Greg Clark; Katherine Fletcher; Stephen Metcalfe; Graham Stringer.

In the absence of the Chair, Aaron Bell took the Chair.

Questions 97-173

Witnesses

I: Jen Gennai, Director, Responsible Innovation, Google; Adrian Joseph, Chief Data and AI Officer, BT Group; and Hugh Milward, General Manager, Corporate, External and Legal Affairs, Microsoft UK.

II: Professor Dame Wendy Hall, Regius Professor of Computer Science, University of Southampton; and Professor Sir Nigel Shadbolt, Professorial Research Fellow in Computer Science and Principal, Jesus College, University of Oxford.

Examination of witnesses

Witnesses: Jen Gennai, Adrian Joseph and Hugh Milward.

Chair: The Science and Technology Committee is in session. I am Aaron Bell, MP for Newcastle-under-Lyme. I am chairing the Committee today in the absence of Greg Clark, our regular Chair. Greg is not only the Chair of this Committee but an excellent Back-Bench MP, and he is taking his private Member's Bill—the Protection from Sex-based Harassment in Public Bill—through Committee today. I know he wants to extend his apologies to you that he can't be here, but it is important legislation and we wish him well with that this morning. He does hope to join us later, depending on how long the Bill Committee takes.

This is the second sitting of our inquiry into artificial intelligence in big tech and UK Government activity. I invite Members here to declare any relevant interests before we start.

Stephen Metcalfe: Thank you, Chair. I declare that I am the co-chair of the all-party parliamentary group on artificial intelligence.

Q97 **Chair:** Excellent. Thank you, Stephen. We have three witnesses for our first panel, representing companies in this space. On the left, we have Hugh Milward, who is the general manager for corporate, external and legal affairs at Microsoft UK. His background is in politics and reputation management. He also chairs the CBI Tech Council and sits on the advisory board of the Institute of Coding. In the centre, we have Jen Gennai. She founded and now leads Google's responsible innovation team, which operationalises Google's AI principles to ensure that its products "have fair and ethical outcomes on individuals and society broadly". She previously led user research in trust and safety, and founded Google's product fairness testing team. On the right, we have Adrian Joseph, who is the chief data and AI officer for BT Group. He was previously a partner at Ernst and Young, leading its AI in financial services activity in the UK and Europe. He has also held cloud, digital, data science and analytics leadership roles at Google. He is a former non-executive director at the Home Office, and sits on the UK Government AI Council. So we have an excellent panel of witnesses here.

I will ask you all individually about your businesses, but to start, I was struck the other day watching a video by one of my favourite YouTubers, Tom Scott. He talks about these sorts of issues quite a lot. He reflects that we seem to be in the same place with AI as we were in with the internet maybe 25 years ago, when we had the first real disrupters such as Napster in that space. We weren't really clear how disruptive it would be, and of course it ended up disrupting millions of jobs and industries around the world. It is a sigmoid curve. Where do you think we are on that curve with AI? Are we really at the beginning, or are we already

quite some way through the AI progress? I will go left to right, and I will ask about your individual companies after that.

Hugh Milward: Thank you for the invitation to be here today. Where are we on the curve? I will answer the question a bit like this. The opportunity of AI is very significant. Over the last week we have seen various instances of AI, including ChatGPT, which is the latest to get the news headlines. Actually, we have had AI around for a while. Where is it on the curve? There is a lot further to go, but the opportunity that AI affords us is tremendous. Chat is obviously exciting, but it is not the entirety of AI. When we look at governance and AI and where AI has got to, we are talking generally about the opportunity of a new industrial revolution.

That curve is really quite long. AI genuinely has the capacity to solve some of society's biggest challenges. It can parse huge amounts of data. It can help doctors, for example, to analyse scans at speeds that are an order of magnitude improvement on what we currently have. It can address climate change issues such as mass flooding and heatwaves, and increase the development of clean energy in a way that we have never seen before. But it can also help to address economic challenges. With the increase in productivity, the growth opportunity there is tremendous. Companies like ours are effectively developing Lego blocks, if you like, that can be built by other companies into solutions that help their own customers to achieve things they could never have imagined before. It is being built into products and services; if we are going to capitalise on this opportunity, and get that curve and really ride it to where it is going to go, we have to make the sorts of investments that companies like ours around the table have been making in digital skills in particular to ensure that we can really leverage this. The UK has this potential around AI to really extend that curve, but we need to get regulation right if it is going to serve society in the way we want it to, but also do so in a responsible way.

Q98 **Chair:** Thank you, Mr Milward. Ms Gennai?

Jen Gennai: I agree with Hugh; AI has been in development for decades, so I probably see where we are in the curve less as a linear increase, and more that it has had its ups and downs. People will have read about things like AI winters, where it got unpopular for a while because people did not see where AI could help to solve some of the real problems and opportunities for economic development, commercial development or otherwise. We are now seeing more of that potential.

Where I do see the maturity being a bit different now is around AI governance. There has been a lot of learning about ensuring that there are guardrails, and considerations about how to develop and deploy AI responsibly. That curve, I would argue, is more mature than it was for the original internet. In terms of the potential, we have not seen all the areas where AI can be helpful from a societal level and a commercial level—that is pretty exciting right now—but I would argue that the governance is where the maturity is actually further along than it was originally in the internet age.

Q99 **Chair:** Thank you. Of course, the purpose of this inquiry is to inform the Government of our opinions about what we need to do with governance. Mr Joseph, what are your opinions on where we are on the curve with AI?

Adrian Joseph: It is hard to tell. We have seen so many advances in the last few years, most recently with generative AI and large language models, of which ChatGPT is a good example. That drive in advances has been underpinned by three things: first, our ability to manage much, much larger datasets with cloud infrastructure, for example; secondly, the compute power required to run some of these models has immensely transformed; and thirdly, there have been huge advances in the techniques themselves. Those three things have come together. It is still such an immature and not fully understood technology, but it is certainly my belief that it will be the most disruptive technology trend in the next 10 years.

Q100 **Chair:** Let us look at some of your individual cases. Mr Milward, what led Microsoft to decide to make its recent multibillion-dollar investment in OpenAI, and what do you think the partnership with OpenAI will bring to your business?

Hugh Milward: What led us there? The opportunity of having an organisation that shared our values as a partner was just tremendous. We felt that having them continue to operate as a sort of independent organisation, supported by us but without the bureaucracy that comes with being part of a big company, was the best model, and that our best opportunity was to support them in where they were going. The opportunity is really tremendous. We can look across the Microsoft suite of services at how we build that kind of large language model and generative AI into Word and Excel, but also into our cloud services—Azure. What we can achieve by doing that with OpenAI in this kind of partnership is really tremendous; it is only limited by the imagination of the people who want to use it.

Q101 **Chair:** ChatGPT has taken us by storm this last couple of months. It has been used to generate speeches in the House of Commons, and to generate lots of content across the internet. Have you seen anything that concerns you from that sort of open adoption over the last six weeks or so?

Hugh Milward: It is important to bear in mind that ChatGPT is one small sliver of an example of AI in use. Let us think about the opportunity for changing the way in which people use search. Search was really defined by the organisation sitting next to me—Google. What they did really set the direction for search in the past, but we think that the opportunity for a different way of providing people with the information they are looking for is also extremely exciting. We are not the only ones looking at that model. It changes a number of aspects of the way in which we go about doing it. It is not generally available—the integration into our service is not generally available. You have to join a waiting list. The latest developments are shared with just a few people. As it becomes a little

more mature—the later adoptions of it—more and more people are able to use it.

We are learning a huge amount as we go. You probably saw about the journalist who spent a bit of time chatting to the service and where that led. We learned so much from that.

Q102 **Chair:** Did it concern you that that was possible? Were you aware that that could have happened?

Hugh Milward: The principles that we have put in place here are iterative. We cannot create something in secret, as it were, and then launch it, because we will get it wrong, and it will not do the things that we want. What we have to do is allow some users to use it and have a process in place where we respond very quickly to what is happening, almost in real time, so that we can address it as we go, really quickly.

The model of governance that we are trying to build, in terms of our principles and the way in which we develop and use AI, is predicated on the idea of very fast iteration. When we get issues like that, which were concerning, we are in position to be able to engineer change extremely rapidly, which is what we were able to do then. We will continue to learn from that experience. It is only by allowing people to use it that we were able to learn. If we had not done that, we would not have been able to learn something.

Q103 **Chair:** From the sphere of politics, are you concerned about political bias and the responses that ChatGPT has been giving to users?

Hugh Milward: It is interesting on aspects of bias. In the principles we have built around the way in which AI will be developed and used, we have bias as one of those core focus areas. How do we ensure that we are not introducing systems that either end up with content that is biased or use sources that are biased? Disinformation is obviously another area of huge concern.

Where would we go in terms of how to regulate for that or solve that problem? We would look at not regulating the AI itself, but regulating the use case of AI, in order to make sure of that. That is the way in which it is deployed and used, in order to try to eliminate some of that bias.

Q104 **Chair:** Because you will be facing some of the same problems that Facebook has experienced, for example, over the advertising of political events. If chat searches are coming back with explicitly political guidance, you are going to face some of the issues that Facebook found.

Hugh Milward: We are. One of the things that we have built into this—the integration of ChatGPT into our service, as opposed to the service provided by OpenAI, which is a separate company—is a reference for the source of each claim. When you put in a query and it gives you something that looks human in response, each statement of opinion or fact has a little number and it references the source. As a user, you have the ability to have a level of transparency about where the information is coming from that you would not have otherwise.

If you think about the deployment for the general public, that is obviously essential. But think too about the deployment within an organisation. For example, if the Government were to integrate this into services, so that it could find the source of material—clauses in Bills, policy across every part of Government—you must be able to have that transparency about where it is coming from, and have that referenceable material. With this kind of service, you get that huge speed, but you also must have that transparency about where it is coming from, so that people are clear.

Q105 **Chair:** I believe the new Department for Science, Innovation and Technology has already had a number of requests about whether they are allowed to use ChatGPT to help produce briefing materials. Do you think ChatGPT or other forms of AI could be beneficial to the machinery of government, as well as commerce more generally?

Hugh Milward: Without a doubt. If it is not that, it is other types of AI. If we think about the challenges that we face in policymaking and in governing, wherever we face a challenge, the opportunity is to be creative about how AI can help to address that challenge. For example, AI can detect anomalies in vast amounts of data. It can parse material. Doctors are using AI to scan for cancers, for example, and increase the accuracy and speed of diagnosis by several days. If we do not integrate that into the way in which the public sector more broadly operates, it is a missed opportunity.

Q106 **Chair:** Thank you. Ms Gennai, how has Google responded to the development of tools such as ChatGPT, and what are some of the other ways Google is using AI in its activities?

Jen Gennai: AI is currently used across most Google products. Everyday ones that people might be familiar with are our Google search product, to get more contextual responses, and Google Translate, which allows for translation for people who do not have access to a lot of digital information because it is mostly in English, to provide economic and educational access. In Gmail, it allows us to remove spam. In our trust and safety efforts, which is one of the areas I have worked in at Google for a number of years, we have used AI to identify potential fraudulent activity online and malicious actor behaviour, as well as the everyday removal of spam from people's inboxes.

For us, AI has been integrated for the past couple of decades in improving our services and providing better consumer products and commercial products. Our cloud commercial side has also integrated AI into its products for our commercial partners. For us, it has been a steady drumbeat of integrating AI while doing ongoing testing. I agree with Hugh that it is about monitoring, testing and phasing releases to ensure that it has been tested internally. As Hugh mentioned, it is really only once it gets into the real world that we can see how users will use it, where they see potential and where they also see potential harms. We learn from that to update our AI.

Q107 **Chair:** I will ask you the same question that I asked Mr Milward: what

have you seen out in the wild that has concerned you about the way in which AI is being used or the responses that AI has been giving its users?

Jen Gennai: Part of our role in responsible innovation is known as adversarial testing, where we try to break a product from different parts of our AI principles. Our AI principles are our ethical charter for the development and deployment of AI. We use each of those principles to try to essentially break the product. We also have a principle around fairness, which is that we will not create or reinforce unfair bias, so we conduct tests to try to break the product from a fairness perspective.

For decades, security teams have done cyber-security through red team testing, and we have done a similar thing with our principles. We do that internally, and then as we release to the wild, we find that a lot of users are essentially augmenting those adversarial tests and providing examples of things we may need to investigate further, including how we train the model to avoid certain use cases and not making it available yet. We are continuing to learn from the use cases, which, through these experiments, allow us to see where certain areas may require guardrails.

Q108 **Chair:** Thank you. Mr Joseph, I know that Graham Stringer has some specific questions about BT, so I want to ask a more general one. It has been suggested that we are in the middle of an arms race. You have two competitors on your right who are very interested in what each other is doing. Do you think we are in an arms race in AI, and if so, what are the implications more generally and specifically for BT and your customers?

Adrian Joseph: We have been in that race for a very long time. This is not new. The big tech companies have been acquiring start-ups and investing in their own expertise for 10, if not 20, years, so yes, it is very much an arms race. It is big tech companies in the US, but in China you also have Alibaba, Baidu and Tencent. Baidu's share price went up 15% yesterday on the back of an announcement that it is going to start building large language models into its search results.

We have seen a lot of activity in acquisitions. We have seen a lot of acquisitions and relationships formed with the big tech companies, whether it is Microsoft with OpenAI, Google with Anthropic, or Amazon, with similar investments in the generative AI start-ups that require the support of large cloud infrastructure to make that a reality and allow those businesses to scale. I do worry about sovereignty, particularly around LLMs. As a member of the AI Council, we have written to the Government and strongly suggested that the UK should have a national investment in LLMs.

Q109 **Chair:** Sorry, in what?

Adrian Joseph: Large language models—the things that underpin ChatGPT, for example, more broadly framed as generative AI. We think that is really important, because there is a risk that we in the UK lose out to the large tech companies, and possibly China, and are left behind.

I think there is a very real risk that unless we begin to leverage and invest and encourage our start-up communities to leverage the great academic institutions we have, to ensure that we have public and private sector all working together, we in the UK could be left behind and at risk in areas of cyber-security, healthcare and so on.

So yes, there is a massive arms race that has been going for some time. The heat has certainly been turned up more recently, and I think it is upon us to do something from a UK perspective to ensure that we are not left behind.

Q110 **Chair:** You have anticipated what I wanted to ask you with respect to your role on the UK's AI Council. A lot of this is being led from California and, as you said, from China. Where are we right now? As you have just said, you are obviously concerned about the risks, but do you think we, the UK, have punched above our weight to date and are you concerned about the future, or are we already falling behind and becoming essentially a rule taker, rather than a rule maker?

Adrian Joseph: No. If you look at the data, we still rank as No. 3 globally in terms of AI capability. Look at the academic institutions, the R&D capability and the start-ups we have—indeed, mine is a classic example, but there are others. BioNTech recently invested over half a billion in one of our key AI start-ups in the UK. So yes, we are.

Part of the role of the AI Council is to support the Government in ensuring that we stay at the forefront of artificial intelligence. Of course, it is looking at the opportunities and how we leverage those, but also at the regulatory environment. I think we have a leading position particularly, in my perspective, around the stance the Government have taken, which is pro-innovation, sector specific and avoiding heavy regulatory infrastructure at the moment. I think that is another area where we certainly could and should lead.

In large language models, we absolutely have the expertise and capability, but we need the convening power of the Government and ecosystems to come together to allow us to unlock that. LLMs, in my view—and I think from an AI counter-perspective—are the biggest area that we really should focus on.

Q111 **Graham Stringer:** Mr Milward, almost everything that is going to happen now badges itself as helping with climate change—you mentioned that. How will AI help with climate change?

Hugh Milward: The opportunities are constrained only by the imagination, really, such as in the development of new materials. We are using AI, for example, in the way in which we are understanding climate change better. We are building effectively a sort of climate computer, which is looking at the data points available around climate, but it is only using the ability to review and understand data at a vast scale that AI can bring that really helps us to draw conclusions about what is happening. If we think about AI as just being chat, we are missing the opportunity for it

to make a difference in the areas where we really want it to make a difference.

Q112 **Graham Stringer:** And has it made a difference in that area? Can you give us an example?

Hugh Milward: As part of this project to build the climate computer, we have partnered with academic institutions all over the world, which are contributing their openly available data to this tool. It is allowing us to do things like harness satellite imagery and connect it to temperature variations over time, so that we can do very practical things. For example, we can identify when a heatwave will arrive somewhere with a high density of tin-roofed houses—there is a particular project in India. What we have seen over time is that a house with a tin roof in a heatwave leads to higher levels of mortality in the area. The tool allows us to make those sorts of predictions, so that people on the ground can do things like put really simple things like hessian sacks on top of the tin roof, which significantly reduces the temperature inside the house, leading to lower levels of mortality at scale. That kind of thing we cannot do without both analysing data and connecting seemingly disparate pieces of data in order to draw conclusions.

Q113 **Graham Stringer:** How is that different from the Met Office being able to predict very accurately for small areas over four or five days?

Hugh Milward: It is almost the same. The Met Office is building the latest version of its supercomputer on Microsoft cloud specifically to be able to benefit from the AI developments that we can bring, so that it may make those kinds of predictions with greater levels of accuracy.

Q114 **Graham Stringer:** How will your customers benefit from AI? If you are saving lives in India, that is a benefit to mankind generally, but how in particular will Microsoft customers benefit?

Hugh Milward: If you take that instance, our customer is in fact the charity or NGO that is using the AI to make a difference on the ground in India; it is not us ourselves. In that respect, in most of what we do—I mentioned this in my opening remarks—we are a bit like a company building Lego blocks, and our customers are the companies that assemble those Lego blocks into tools, services, products or whatever it might be for their own users. The AI that we build can be used by those companies to create new IP, new ideas or new companies that will deliver for their customers.

Q115 **Graham Stringer:** Are there any other exciting and interesting activities that you want to tell us about? I don't particularly want to give you a platform to advertise Microsoft, but if you want to tell us about anything, now is the opportunity.

Hugh Milward: We are working on a really interesting project through our research facility in Cambridge, for example. We are using AI for a number of children who are blind or partially sighted to be able to navigate the environment around them. They can walk around corridors, and in fact the AI is now able to recognise the people whom they interact with

regularly, so that they can have the sort of interaction with the world around them that their better-sighted friends and schoolmates are able to have. It is transformational, but it is the kind of ability that we could not really have without the advances that we have been making in AI.

Q116 **Graham Stringer:** Thank you. Jen, are ChatGPT and Bard the main areas where you envisage AI being a part and where it will expand in the future, or is it other areas?

Jen Gennai: It is one of many areas. There are so many examples of where AI is already used in products, as I mentioned. The obvious example is climate change—which you have already asked about—where it could be helpful. We are not currently using generative AI in some of our climate change initiatives, but some are already becoming increasingly important.

You may have read that some of our climate change efforts have focused internally around our data centres, ensuring that we are using less energy for cooling—less water usage. Water is one of the most critical issues of our time, so it is not just about water usage in Google's data centres but about how we support that globally, such as thinking about new agricultural processes.

Again, through AI, there is the complex chemical nature of some fertilisers. How do we reduce the energy required to produce them, and the water required to grow things, to increase yields in an increasingly warm, water-constrained world?

But it is not just about what we can do now to manage climate change. It is also about adaptation, with things such as energy usage and understanding things like nuclear energy. We know, across the globe, that some countries are increasing their investment into nuclear energy, but we also know that it is a massive capital investment, with some risks about where waste is stored.

AI can help us to analyse ways, through nuclear modulated reactors, to launch them sooner, and find out how to manage the waste in a safer way. It can also help, as you mentioned, with thinking about understanding new chemical ways of bringing energy together, reducing the waste and increasing that. For us, AI can be used across handling climate change, adapting to climate change, and trying to reduce its impacts.

Similarly, as you mentioned our commercial partners, the use cases that they have for AI range from improving their own internal operations, through faster digitisation and summarisation of documentation to supporting their customers better through customer support channels; we know that emails or phone lines may not be appropriate for individual needs. We found that AI can provide more contextual support for customers across any company's needs—any Government needs—to allow people to get their questions answered in a faster way. We see AI across a range of that. Generative AI is one pillar across the AI spectrum that will

help us to address those needs, but we have so many other areas continuing to be invested in.

Q117 **Graham Stringer:** Thank you. Adrian, can you tell us how BT customers will benefit from AI?

Adrian Joseph: I didn't know before I joined BT three years ago that we were the No. 1 patent filer in the UK for artificial intelligence. I didn't know that we had 25 years of R&D investment in AI. I didn't know that we had relationships with 20 of the top universities in the UK, and globally, to help drive these things forward.

So how have we used our expertise, over the years, to help customers? Well, there are a number of ways, but I will share a couple of them. We are looking at things such as wi-fi analytics at the moment—for example, the hub in your home. We can detect signals using analytics and AI, perhaps for an elderly person who may have fallen down and not got up. We can sense things like that.

You mentioned climate change. One of the models that we worked on was around workforce optimisation. We have a very large field force of people going out and doing maintenance and repairs. Through some of the AI techniques that we have used, we have managed to get 400,000 more visits and brought travel time down by 17%. That is another 39,000 tonnes of CO_2 emissions that have been taken out of the system.

We use it in other ways to think about nuisance calls. We have a fairly significant cyber-security business, so we look at ways in which we can detect threats. We have seen a spike of about 400% in cyber-risk in the past two years, so we are using a number of techniques in our own products to detect and address those threats before they become realities for customers.

We look at our network; we look at anomaly detection, which is another core area for AI to figure out. Are there any issues that we can be aware of and rectify before they hit customers? And, of course, we think about other ways in which we might improve the customer experience—by connecting to the best agent for that particular customer, for example.

There are a number of ways in which we think we can either improve the efficiency of our business, and pass that through, with more value to customers, or get better at the recommendations that we make to them, or get rid of things such as nuisance calls and improve overall customer service for them.

Q118 **Graham Stringer:** You said that you wanted to deliver AI-led operations by 2025. Why 2025?

Adrian Joseph: To deliver what, sorry?

Graham Stringer: You said that you wanted to deliver AI-led operations by 2025. Why 2025 and what will they look like, beyond what you have just told us?

Adrian Joseph: Those are all good examples. Another good example would be the AI accelerator that we have just launched. It is an MR platform that allows us to scale what we do across our business in AI operations. It drives a lot of consistency, and it drives a lot of speed. Through this new platform, we have been able to reduce the time that it takes for getting AI into production. We were at about 180 days at the start of last year; we are now down to about 20 days.

My next improvement is in speed. It is not just about speed, of course. It is also about safety, so we have lot of components built in around explainability. We have a framework called Responsible AI, which looks at four key principles around fairness, openness, accountability and so on. We need to do both of these things. Our simply stated mission, which is not so easy to deliver, is to radically and safely drive pounds—as in sterling—per petabyte. We have a lot of petabytes of data, and we have a lot of AI assets that have yet to be fully turned into value for us and for society. Our goal is to do both of those things to support the UK as well, of course. Through things like that, we believe that we will be able to deliver by 2025. I would hope it is earlier than that, but I am happy enough to put 2025 in the public domain.

Q119 **Graham Stringer:** This is my last question, and I thank all three witnesses. It is a different kind of question. You have been through the potential and real benefits of AI. I get worried about AI; it is difficult to understand, and I have a scientific background. I get worried about the size of your companies and your being very powerful in democracies, and I get worried about genocidal regimes such as China. I know Google barely operates in China, but should I be very worried and concerned if these AI products are released in China to a pretty appalling regime? Hugh, maybe you could respond to that.

Hugh Milward: We would equally be worried about that. It comes back to some of the discussion that we had a little bit earlier about the emphasis on UK sovereignty and AI. Actually, in some respects we might want to look at the values of the companies that are developing AI and look at the use cases of AI. From a regulatory perspective, I think we are going to benefit most in the UK by ensuring that we support the UK's opportunity in the development of AI, and Adrian has outlined that very clearly. There is a great opportunity—we have a great advantage here—but we cannot hold back the development of AI that is being developed in other countries where there is a high set of ethical principles and values as a counterbalance to some other countries where that is not the case. We have talked about China, and that is one example of an area where we have to stay ahead in "the West" with values that are effectively pro-democracy. We have to allow that to flourish if we are going to ensure that those that do not have a counterbalance.

The other thing is to make sure that, as we regulate AI, we are thinking hard about the regulation of uses of AI rather than the AI itself. If we do that, then AI, irrespective of where it is developed—it could be developed in a regime that we do not necessarily agree with. If we are regulating its use, then that AI, in its use in the UK, has to abide by a set of principles.

Where we have no ability to regulate the AI itself and where it is developed in China, we can regulate how it is used in the UK. That allows us then to worry less about where it is developed, and worry more about how it is being used, irrespective of where it is developed.

Q120 **Graham Stringer:** Are you not worried about Microsoft AI products being used to repress the Uyghurs in west China?

Hugh Milward: Without a doubt. The way we develop our technology and license our technology includes restrictions on use. We can define when a customer is allowed to use it or not, and we have. We have made decisions within the company.

Take a technology that is "dual use", like facial recognition. You can use facial recognition technology—the same technology—to recognise a cancer in a scan. You can use the same technology to recognise a missing child. You can use the same technology to identify a military target. You can use the same technology to identify citizens who are unlicensed in a regime, who you might want to haul in. That is the same technology being used for very different purposes.

If we were to restrict the technology itself, we would not get the benefit from the cancer scans in order to solve the problem with its use in a surveillance society. If we regulate for its use rather than for its development, we can quite comfortably say, "No, we are not going to license this technology for you, and we are not going to license that use of this technology," even if we will license a different use of that technology.

Jen Gennai: I will add a bit to that. I definitely agree around ensuring that the UK is leading with innovation—ensuring smart regulation in that regard. But I also support your efforts to inform global standards. Setting a global expectation through organisations like ISO, IEEE, OECD or G7 gives you a chance to inform what the global standards should be and raise the expectations across the globe in that area.

From Google's perspective, I agree with the use cases, and we have restrictions around what the use cases will be, at the deployment level. But our AI principles also take into account as we develop and deploy—I mentioned our principles in general. Some people may be familiar with our seven aspirational principles, which are around things like be accountable to people, and provide social benefit. Maybe not as many people are familiar with our four guardrails, which are applications which we will not pursue. We don't develop and deploy technology that overall causes harm, that supports surveillance, that violates international norms, that violates human rights or that can be used in weapons or a weapons system.

Even as we develop technology, we are not just waiting for how it could be used or misused; we also ensure that our technology is not developed in any way that would support any of those use cases. Again, throughout the life cycle of the AI development and deployment, Google is trying to ensure that it could not be used in ways that we do not support.

Adrian Joseph: There is a broader piece around how we think about this. There are three concerns around privacy, around bias—bias in three ways: bias in the data, in the algorithms and in the individuals that are creating some of the algorithms. I would argue that all three of those are equally important. If we had better diversity in the individuals that are creating these algorithmic capabilities, we would have seen many fewer negative consequences, whether that is gender-biased algorithms or racist algorithms.

Ethics is another area that we look at very hard. It is not just about whether you can do something, but whether you should. When I was on the board of the Home Office, I was part of a data ethics group, where we spent a lot of time thinking about ways in which data was brought in, and whether it was being used for its intended purpose.

The third area I worry about is uncontrolled automation. Our principle around AI is around a human plus algorithm always outperforming human or algorithm—so the idea of having a human in the loop. I do worry about your wider point—about the negative ways in which AI could be used. Your example of China is a good one.

Chair: Thank you. It has been fascinating so far. We are running a little bit behind schedule, so if we could try to keep the answers a little bit shorter in the second half—though I would not fault any of the answers so far.

Q121 **Dawn Butler:** Thank you all for your evidence today. I want to come on to bias, where you have just ended, Mr Joseph, but first I want to ask this. You said that we are number three with regard to AI. What do you think we have done to deserve being number three?

Adrian Joseph: A tremendous amount. Look at the educational system and academic system that we have got. Look at the number of start-ups we have created that have been spotted by many of the big tech companies—too many of them, in my view, have been acquired.

Look at the work that many universities are doing at the moment, whether it is Oxford, Cambridge or the London triangle, or universities such as Manchester, which have now formed a grouping to figure out how to allow entrepreneurism to flourish. They are taking equity stakes in some of the new ideas that are coming about. That has been a really strong ecosystem.

Some of our own companies are doing great work. I have mentioned some of the things that we have been doing at BT. We have been doing this for more than 25 years, with the relationships that we have formed. The ecosystems that we have, and how we work with some of the technology companies and other partners, are really important.

Healthcare is another area where we have done very impressive work. If I think about the regulation that I mentioned earlier, we are and can continue to lead in the area of AI regulation and ethical use of AI. Those

are some examples that I would give. But we are at risk of dropping down in the rankings.

Q122 **Dawn Butler:** I understand; we are ranked at number three. Can I talk about the risks and potential harms? This is for all of you. We can focus on the benefits of AI, but if we do not now focus on the risks, and mitigating those risks, further down the line we are going to find ourselves in a situation that we cannot get out of, and we cannot explain how we got there in the first place. What specific things are your companies doing to mitigate the risks of AI? I will start with Mr Joseph.

Adrian Joseph: Thank you for another great question. There are many things. We have a responsible AI framework, which I mentioned, and we are keeping scores around that. It has got to be for good; it has to be transparent; it has to be fair, and so on. For any of the new AI capabilities that we are building, we have a steering group that is across responsible tech. Our privacy lawyers, our legal teams and our technology teams all come together to look at the implications and potential risks that we need to be aware of.

Q123 **Dawn Butler:** How do you disrupt that? How do you not get comfortable with that? It is okay to have a list and go through that, but how do you ensure that you disrupt that at regular intervals?

Adrian Joseph: That is the starting point. I mentioned the AI accelerator platform. Built into that are capabilities around explainability, transparency and automation of looking at things like model drift and data biases. We have both a framework and a process, but also a very heavily technology enabled platform, as well as human oversight, that allows us, on an ongoing basis, to detect, address and remediate those types of risk. It is an area where we spend a lot of time. Of course, bias is a huge area, as I mentioned before.

Q124 **Dawn Butler:** This is my last question to you. How do people respond when you talk about bias in the systems in the way that you do? Is there an understanding of the response or is there sometimes a denial?

Adrian Joseph: Of the first two that I mentioned, the bias of data is a very obvious one—if you have rubbish in, you will have rubbish out. The bias of algorithms, and being aware of model drift, is, I think, also fairly well understood. We are seeing more automation built into platforms, to enable us to deal with that.

The third bias is a harder one. If you look at diversity and inclusion in the UK, only one in 10 data scientists is female. That is not good enough, so one of the things that we spend a lot of time thinking about is how we ensure that our teams, teams that are working for us, are much more representative of the audiences and the customers that we serve. One of the things that I did very early on was to hire a female head of AI. She was hired because she was the best person, not because she was female, but now 30% of our data scientists on AI teams are female and over 25% are from black, Asian or minority ethnic groups. As I said before, I think that if we had had more types of diversity reflected in this sphere of

technology, we wouldn't have seen as many of the examples as we have done. That's the harder area, I think.

Jen Gennai: We think about managing AI risks in three main areas. One is understanding and values; the second is around monitoring and managing; and the third is infrastructure and scale. What I mean by "understanding and values" is that we want everyone in Google to understand that developing AI responsibly is their job; it is not just my team's job. We are the central team; we are accountable for the company, but we train people in Google to understand these AI principles, adopt them day to day and build a culture of responsibility every day. We do that in many different ways and not just through formal education plans. Certain roles require trainings to take in tech ethics, our AI principles and our governance structure. But there is also more accessible training. To your point, some people may feel that even by hosting an ethics training, we are saying that they are being unethical. That is not what we are trying to do at all. We need to make it accessible and show that this is their responsibility. The very deliberate name of our team is "Responsible Innovation", of the ethics team, because "ethics" can imply that there is a personal judgment. What we are saying is that everyone has responsibilities to follow in Google to ensure that they are developing AI responsibly. So that is our first bucket: education, culture change and accountability, essentially, across the whole company.

The second one, around monitoring and managing, is my team's responsibility. We are a central governance team who provide review processes for technology as it is being developed and throughout the lifecycle. We conduct tests on the data, on the model and then on the output. I mentioned some of those tests earlier. We do adversarial testing—human adversarial testing as well as technical adversarial testing—to try to see where the product may have anomalies and may introduce risks throughout the lifecycle, and then we work with teams to try to prevent that. Also to your point, to avoid getting complacent or being too brittle, we don't use formal checklists. We do have review considerations. So there is a series of questions, certainly. There is a series of tools that teams are expected to use. But they are adapted to the use cases and the needs of the technology end user.

Q125 **Dawn Butler:** How do you address the potential for generative AI to spew out misinformation? The problem is that it can predict the next word, but it doesn't understand the sentence.

Jen Gennai: I agree that artificial intelligence is a bit of a misnomer at the moment, because it has to be trained in such a way as to be intelligent, to understand the actual context of the understanding. Part of what we have done, although we have applied similar governance structures to our generative AI development work—we have learned, again, that there are differences in how much work is required to work on factuality and the notion, which you may have heard about, of hallucination in generative AI, where AI does not know what it is saying in return to a perfectly benign question and then it starts to hallucinate. We are continually doing tests on that, because it is an unsettled question at

the moment. So although we are following a similar process, we are finding new areas where we have to go deeper into our research. We have hundreds of researchers specifically focused in this area. We have been adapting our processes to try to understand where generative AI may be raising factuality and fairness risks more than accountability risks or other risks. So we are trying to be adaptive while following—

Q126 **Dawn Butler:** How can we legislate to make sure that there is enough care in the process around that?

Jen Gennai: That is an excellent question. We try to show our work through areas like explainability and accountability. One of our principles is to be accountable to people, and that is not just to regulators and stakeholders, to be clear; it is also to our end users and our commercial partners. We try to provide explanations of what the AI is supposed to do and what it is not supposed to do. You may have heard of our artefacts known as data cards and model cards. If you think about a nutrition label on a bar or a cereal that tells you what is in them, that is what our model cards and data cards try to do; they tell people in real words—not a whole algorithm—what the model is trying to do.

We share that through those model cards, but we also share more user-readable explanations: "Here's what this model is trying to do, and here's what you should expect in return." We are working through whether those explanations happen in real time, as you are using the product, or as you sign up for the product—I believe Microsoft is also doing this—which involves saying, in this instance with generative AI, "This is an experiment. You may see things that are not up to par, and that is what we need to learn from you," so that they are clear about that and we are transparent about the maturity of the product. That is how we try to keep people informed.

Q127 **Dawn Butler:** How do you build out bias in the system?

Jen Gennai: We try to take a phased approach across the data, the model and then the end product. In the data, we try to ensure that we test for potential bias in the data, because unfortunately, humans are biased; we all bring our own personal lived experiences and bias. That is where AI has learned from existing data from humans, so we have to go back in and try to see where there are likely exclusions of certain sub-groups. As I mentioned earlier, a lot of language online is English. A lot of languages are not represented there, so we try to get more language data and representative data to build our datasets.

We then move into how the model is being built and how we identify whether there are potential differences in sub-groups. Again, we may see that women are not represented in the text. How do we get more data related to that and build up the model to learn? My team then comes into play for testing, as the product is ready to go live. We work with civil rights groups and human rights groups, as well as conducting user research and tests. Once the product is almost ready to launch, we are testing it to see who may be excluded from this experience and who does

not have as positive an experience as someone else, and then we work with the product teams to fix it.

Q128 **Dawn Butler:** Do we have to accept that we cannot ever build out bias in the system?

Jen Gennai: It is very hard to say that there will ever be fully fair products. They should always be fairer than they would have been without these sorts of tests and balances, and it should involve constant monitoring. Once we launch a product, it is not hands-off; it is continuing to iterate and learn from our own technical infrastructure that will identify anomalies from a fairness perspective. We also make those tools available to other people. We have tools like fairness indicators that essentially indicate how fair your model is. Google built that for internal uses, but we have now made it available for our commercial partners through TensorFlow and other ways. There should be constant monitoring, but because humans will never be unbiased, we have to check that they are not always inserting bias into the models and products.

Hugh Milward: I am not going to repeat what has been said, as I appreciate that we are short of time. AI systems are not going to be able to correct the biases in society, but we want to make sure that they do not amplify the biases in society. You will see bias depending on where you stand. For some of the population, the BBC is a left-wing organisation. For some of the population, it is a right-wing organisation. The important thing is to make sure that we are not amplifying the biases that are there.

The way in which we think about risks and harms is around the design of the system and the use of the system. We want to try to avoid the situation that we have in some respects with social media where, however many content moderators we might bring to a system that is causing some harms, we are never going to solve the problem unless we solve the problem at the algorithmic level. It takes both: it takes the use of the algorithm, but we also have to make sure that we use the principles we have at our heart in the building of the algorithmic system.

Q129 **Dawn Butler:** Is there a case to say that we cannot allow organisations to self-regulate? The idea was that social media companies will self-regulate and put all these processes in place, but actually, it does not work.

Hugh Milward: Certainly with AI, the need is to regulate the use rather than the AI itself.

Q130 **Dawn Butler:** Yes; you made that point very well. There were a couple of things. The example that you cited on AI about tin roofs concerned me a little bit. I don't think you need AI to let you know that if you have got a tin roof, you are going to be hot inside. Probably the better solution would be to have solar panels on top. It worried me a little; in some situations, some organisations are taking the data to match what they are doing, rather than what they are doing producing the data.

Hugh Milward: The ability to understand data is what AI can help you with. If you are an NGO working in a shanty town in India, knowing where to deploy your limited resources at certain times of the month or the year or even the day, makes the difference between some communities being able to thrive or not. You could say you could do that with paper and pen, but it is quicker with the computer.

Q131 **Dawn Butler:** Yes, I understand that, with numbers and so on.

You talked about advanced testing and how you choose certain people. Who do you choose? What is the process for the people that you choose?

Hugh Milward: We start internally. There is rigorous testing internally with engineers. There is a whole ecosystem built around people volunteering internally to test our products and services. Then people can opt in. They sign up. Then we have limited acceptances of people who sign up, in small numbers, and then we gradually increase the numbers who can use it.

Q132 **Dawn Butler:** Is it always internal?

Hugh Milward: No, absolutely not. I want to reiterate that OpenAI is a separate company from us. They are not owned by us. We haven't acquired them or anything like that. They run their own model of testing. But the integration of ChatGPT into Bing, which is the Microsoft search engine—we have been very limited in who is part of that ring, the increasingly-sized ring, of people using it.

Q133 **Dawn Butler:** Do you make sure that the people that you use for your advanced testing are diverse?

Hugh Milward: That is a good question. I actually don't know the answer to that.

Chair: Perhaps you could write to the Committee on that.

Q134 **Stephen Metcalfe:** At the beginning, the Chairman asked where we were on the curve of development. Is the reality not that we have no idea? If you go back to the '40s and '50s and the number of computers that were predicted to be needed in the world—now we have a chip in every lightbulb, in every car; we would never have predicted that. So we don't know.

What I really want to ask about is your view on the future. What we have learned from the last 100 years or so is that sci-fi becomes sci-fact, certainly where with the tech, there is a visible path that doesn't require completely new discoveries. Where do you see the future of this leading? Should we be worried about where that future may take us? What do we do about that, and when should we be worried?

Hugh Milward: I am not worried about the future. I am actually pretty excited about what this can do. That is largely because the principle that we have worked on is that AI is in effect a co-pilot, not an autopilot. Its job is to augment the things that human beings are doing, rather than to

replace the things that human beings are doing, and to really allow humans to be more human, in some respects.

I also think that if we get the balance of regulation right, particularly here in the UK, and maybe don't go down a route that some other countries have gone down—I am looking particularly at the considerations taking place in Brussels about what they are doing. We should not repeat that because it will both stifle innovation—we really want to drive innovation in the UK—and because it also doesn't solve the problems that they think they are going to solve for.

We have an opportunity in the UK to nurture innovation that means that AI can be effectively a rocket booster on the economy, on growth and jobs and that sort of thing. But at the same time, if we get it right, we can do so in a way that makes sure that those guardrails around the development of AI are in place, which means that people can be more confident and more comfortable about AI serving humans, as opposed to the other way round.

Stephen Metcalfe: Thank you. Jen?

Jen Gennai: I am equally excited about the AI opportunities. We talked through a number of them earlier around climate change, health and being able to support that, but also everyday users, like I mentioned. Coming from user research as my focus, there is capacity to improve education and economic opportunities for people across the globe who may not have had that without some of the AI technologies and the facilitation of things like Translate. But my job is also to be worried, so as the responsible innovation lead, it is important, as I say, not just to have processes and policies that we follow—they are represented in both global standards and smart regulation—but to ensure that there are checks and balances. It is important to have those governance capabilities as well. I would have it with a cautious yes: as long as there are guardrails, I am very excited about the technology, because I am hired to be the one who worries.

Stephen Metcalfe: Adrian?

Adrian Joseph: Uncomfortably excited. I am excited about the opportunity ahead of us. Actually, the technology capability that we have now is, in my view, massively under-exploited, in part because I think a lot of our business leaders and others are not fully aware of just how significant an opportunity we have at our fingertips at the moment. As to the future, I agree with you that we just don't know. It feels like we are at the very early stages. Six or 12 months ago, would we have been talking about generative AIs as much as we are at the moment? Both have a dual purpose.

I am excited about some of the things about the future of work. When I think about the impact on work, I think about three things. First, I think pretty much every job will be augmented by AI capability. A lot of the work that we do was not designed for humans, in my view, so that is an exciting prospect. I am excited about the new jobs that we are going to

see coming up, whether it is ethicists, creative profilers and so on—new things that we have not even thought about. I am uncomfortable about things like equality, skills and the UK potentially being left behind—particularly in the area of generative AI, which I mentioned before—around ethics, dual purpose technology and just how scary that can be, both on the upsides and the downsides.

Q135 **Stephen Metcalfe:** Thank you for those answers. With all due respect—which obviously always means that there is none—all those answers are within a visible horizon. My concern is that that visible horizon is fairly short term. I want to be able to see over that horizon and whether or not there are things that we should be doing now that will stop us turning round in 50 years' time and thinking, "We didn't intend that to be the outcome." You talked about having a human in the loop, but an AI might decide that, actually, we are the problem and that it should remove us from that loop because we are hampering its ability to do the task that we set it. Are there really no concerns or no thinking going on about making sure that that does not happen? If there is some thinking, please can we share that so that we as legislators and regulators—or potential regulators—can know what we should be putting out as guidance?

Hugh Milward: I think there is a really useful model in Mary Warnock, who effectively led a commission that looked at a set of principles and rules around human embryology and fertilisation as an ethicist and as an expert. What she did—I think it was in '89 or thereabouts—was create a set of principles that are as relevant today as they were then, and they were created at a point where some of the developments in science could not have been foreseen at the time that they were developed. They were a set of principles that could sustain and guide the development of the technology, without regulating it out of existence and without crushing the opportunity for innovation. I think that is a model that we want to look at for the regulation of AI.

It worked because it is principles-based, rather than trying to solve very specific technological problems. If we take an approach in the UK that looks again at the principles and is sure that these are principles that will sustain, and wherever the technology goes they can still apply, we will be in a really good place to get that generation of innovation at the same time as guiding society. It is also important to check back in that we are going in the direction that we want to go in. As this is fast moving, we need to have a regular review: "Are they doing the things that we want? Do they need to be adjusted?" There is a useful model there.

Jen Gennai: I'd agree with a principles-based approach that allows for support of innovation while ensuring guardrails, but let us be a bit more tactical—this goes back to Ms Butler's point—by including diverse voices in considering potential risks and harms. We have definitely leveraged experts external to our company who are thinking about these types of issues, from civil rights groups to human rights groups and futurists, who try to challenge our thinking and get us thinking in a bit more of a future-looking way. We also do exercises known as horizon scanning and scenario planning, which take us out to future scenarios to stress test whether they

are likely—whether we are on the path to that or whether we have a chance to guide it. They are two of the main approaches in addition to a principles-based approach that I would suggest keeping in mind.

Stephen Metcalfe: Thank you. I'll wrap up after we have heard from Adrian.

Adrian Joseph: I agree with the comments that have been made, and with the way in which the UK approach to regulation is developing. That is sector-specific, so the FCA—or whatever the particular regulatory group is—is involved. We also need horizontal perspective, to look at the overarching areas around bias, for example. We will have to have perhaps even more international collaboration, and sharing of issues and concerns, than we have at the moment. We do need to hold ourselves and big tech—or all tech—responsible for outcomes, and to be very clear about the safeguards that should and could be put in place to allow that to happen.

Q136 **Chair:** I want to run briefly through what we will do about governance here, but first I just have one specific question to Mr Milward about ChatGPT and its integration with Microsoft, and privacy. Can Microsoft workers read the requests that users make of ChatGPT?

Hugh Milward: Let me talk about the way in which we design the model. Take a company like Vodafone, which built the ChatGPT functionality into its call handling. Effectively, that process uses voice recognition to turn it into text and the functionality around the chat to understand it—

Chair: And then a human takes it over, yes.

Hugh Milward: Vodafone's data—all that deployment—cannot be seen by, for example, Lloyds, another customer that is using it. It is completely independent. They do not learn from each other. The branches are entirely separate. There is no kind of Vodafone customer data that is available to Lloyds—

Q137 **Chair:** I did not mean so much from a customer data perspective. If somebody is talking to Bing about something in detail, obviously that information will be stored by Microsoft, but can people go and look at that, or is it secure and encrypted so that individual workers at Microsoft cannot read it?

Hugh Milward: Search isn't encrypted.

Chair: So individual workers can read the search history of individual users.

Hugh Milward: If you take a function like a search engine and use it to try to find some answers, that is not an encrypted or secure service.

Chair: Understood; thank you. Katherine Fletcher will take us through regulation, and then there may be one more question from Stephen at the end.

Q138 **Katherine Fletcher:** Thank you all for coming; it is appreciated.

This is an exciting field and area. What strikes me from listening to you, and to the previous evidence sessions we have had, is that we are describing a technology in very broad brushstrokes because it is so embryonic. Some bits of the technology are a Model T Ford—me and my spanner could probably take it apart and put it back together again—and some are a McLaren F1 or a Bugatti Veyron or whatever. Interestingly, we have a step after that, which is the intelligence side of the artificial.

I want to home in not necessarily on the day-to-day algorithms in the near term, but on the very smart ones—the long-term planning algorithms—and to test some of the regulatory frameworks in those areas. First of all, what are your expectations of different regulations? Are you expecting to regulate the Model T and the Veyron and the new clever thing in the same way? Let us try to be brief, if that's okay.

Adrian Joseph: Back to the data, the input and the output; I think that model applies to any of the models, whether it is the T Ford, the McLaren 765 or—

Q139 **Katherine Fletcher:** Do you have the expectations you need in place for a good regulatory framework?

Adrian Joseph: From an internal perspective, we believe we do, but there are models where we do not fully understand all the machinations, whether those are in an RNN or a CNN—a neural network. In areas of generative AI, we do not fully understand how some models develop.

Q140 **Katherine Fletcher:** I am trying to give you an opportunity to comment on the record about how the Government and society are regulating you at the moment, and whether their future plans are within expectations or not.

Adrian Joseph: Both from an AI Council perspective and from where I sit in BT, I think the sector-specific and pro-innovation plans that are in place are the right models for now.

Katherine Fletcher: Okay. Well, maybe I'll come back to the future. Jen?

Jen Gennai: I agree that there is a lot of existing regulation that, even internally, our AI principles are very aligned with. For example, privacy is one of our principles for responsible AI development, and there is existing privacy regulation. Earlier, we were talking about fairness and anti-bias; there is also anti-discriminatory regulation, and we have learned from a lot of that and applied it to our technology. For us, there is a lot of existing regulation that we are trying to align with, even if it has not yet launched in many regions specific to AI, because AI is just the underlying technology.

To your point about the car, the question should be, "What is the car allowed to do versus how it was particularly done?" AI is just the "how" it is being done, but it is the "what" that we want to consider. We agree with that multisectoral and principles-based approach. There is a lot of existing regulation, and in other areas is about thinking, "What's unique about AI?". Internally, we have tried to ensure that areas around explainability

and accountability are a bit more different, because of the complexity of AI versus a hardcoded line of text. There are some differences, but there is a lot of existing regulation that we are applying.

Katherine Fletcher: It is interesting that the existing legalities are being factored in. Would you add anything, Hugh?

Hugh Milward: Only that I would say that if we regulate in the right way—I think we are currently getting it right in the UK—there will be no difference between regulating between the Bugatti Veyron and the Ford Model T, and every car in between—and, in fact, any car that is invented beyond that is more super than the Veyron. That is, it is the use of the AI rather than the AI itself that we are regulating.

Q141 **Katherine Fletcher:** That's very helpful. Let me pick up on the one thing that we have heard previously does not sit within that car analogy: the long-term planning algorithms that will be smarter than us. I am talking about the algorithms that will be set a mission and could, in theory, decide that the best way to achieve that mission is to remove all cows from the planet or to remove all humans from the planet—that kind of very long-term information. I am interested in testing whether we can set up starting rule sets for these long-term planning algorithms.

I am a biologist by training, so bear with me on this analogy. The 4 billion years of life plays game theory, effectively, because it competes for its existence, food resources to fuel it and resources to allow it to reproduce in the next generation effectively, whether that is through sexual or asexual selection. We are about to use technology to create something that has an interest in surviving, growing and thriving that we cannot necessarily read—these long-term planning algorithms.

I am interested in your thoughts about whether we should only be turning these things on if we can give it a sense of its own existential threat or, to put it extremely simplistically, we can turn it off. We know how to turn a Model Ford or a Bugatti Veyron off, but, short of pulling every server out of the wall in the globe, do we know how to stop a very clever, well-fuelled future planning algorithm that has been running a long time? I appreciate that it is a very small subsection of a very exciting field, but I am interested in your thoughts. Is there a role for regulators to play, to differentiate AIs?

Hugh Milward: I may not be the best person to answer this. I do not know what a long-term planning algorithm is; it is certainly not something we are developing at Microsoft. The ability to have in place a set of principles that defines not just the use but also the way in which the algorithm operates, and getting both of them right, is incredibly important. I used the social media analogy earlier. It is the design of the algorithm as well as the use of the algorithm that is really important. If you are going to build safeguards in, I genuinely think that Parliament should probably solve for the most common issues as well as having an eye to the most extreme—

Katherine Fletcher: Low likelihood but high risk.

Hugh Milward: That is right. At the same time, you should have a view as to what the extreme circumstances might be and have in place a mechanism or principles for making sure that does not happen. The idea of turning something off is always your option of last resort but, again, I do not know what a long-term planning algorithm is.

Katherine Fletcher: We have heard evidence previously that it would not be possible if they were that smart.

Hugh Milward: To be honest, I do not know who would be developing this in a company like ours.

Katherine Fletcher: I accept that there is perhaps not a commercial interest to it, but the application of technology for both positive and nefarious purposes is the story of humankind. Jen, what is your response to my longer question?

Jen Gennai: I agree with thinking about how this technology is developed and centring the fact that humans have to set the objectives of AI technology. AI does not learn by itself where it is going: humans set the objective function of AI technology, and code it to learn from a series of inputs and then output to that objective function. The smartness of AI is to do that very effectively and in a pretty complex and very fast way, as Hugh mentioned, but humans design the technology and set the objective function. They are the ones who identify where the data source comes from. They also have control over the monitoring of that algorithm and beyond. It would have to be acknowledged that humans would be training it to do something where there may be a concern that it would get out of control. An AI algorithm is built, developed, tested and monitored by humans, so I would maybe not share the same concern that it could get out of control. A human has decided to design an AI to do that.

Katherine Fletcher: That is very interesting. Adrian, do you have anything to add?

Adrian Joseph: I am not aware of us doing any of these long-range models either, but to your question about things that we are thinking about from a regulatory perspective, in addition to the UK's position already, one would be creating a framework of risk and being really clear about what is meant by real and unacceptable risk. Your example may fall into that classification. Cross-sectoral principles and a compendium of use cases would also be really helpful. I mentioned international standards and staying really closely connected to the international community.

The final point I will make is around the skills that regulators need to have. There is work for all of us to do in terms of skills uplift, but I would think really hard about the types of skill profiles that will be required in the FCA, Ofcom and other regulatory environments to potentially avert the long-term risks that you are thinking about.

Q142 **Katherine Fletcher:** The example that's been put to us is an entirely reasonable one: how do you cure climate change? You let a very clever

algorithm go off and it realises that cows are a major contributor to climate change, so it decides that we need to get rid of cows. Because it is in the black box bit, we are not aware of the fact that because it controls the feed-cow manufacturers—because they are all on the internet of things in 50 years' time—somehow, by some nefarious purpose, the cows are nixed and we don't know about it because it is a black box decision. My interest is in how to make sure there is a consciousness of the system that the programme sits within. It needs to know that it relies on electricity and that if it poisons the planet, it is also knackered. Do you see what I am getting at?

Adrian Joseph: Yes. I would go back to Jen's point about the remit and the context in which the algorithm is directed. It is about having a human in the loop, ongoing checks and balances, and adversarial models that are checking to see whether the right outcomes—

Katherine Fletcher: So even though it is black box, a human could monitor actions, for example.

Adrian Joseph: I think so. I mean, the ultimate black box is our brain.

Katherine Fletcher: It's an interesting topic, but I am very conscious of the time. It has been helpful to hear your perspective on how likely that is within our lifetime horizon.

Chair: Thank you, Katherine. I am very conscious of the time—we are keeping the next panel waiting—but let us have one final question from Mr Metcalfe, and quick answers.

Q143 **Stephen Metcalfe:** It is a small question but there are two parts to it. First, who should we look at internationally for some help and assistance, either to avoid doing what they have done or to look at how they are doing it well? Secondly, as you have the platform of a Select Committee, what one request would you make to the Government to support the work you are doing in the sector?

Hugh Milward: I would look to the EU for a model of how not to do it. I actually think that the UK is probably taking the most progressive steps that we have seen anywhere, so it would be very sensible not to break what we have here that is not broken. We should be doubling down on the approach we are currently taking.

I do think that in the UK we risk creating a regulatory environment that is business unfriendly. Certainly from the international-investor perspective, we increasingly see the UK as a tough regulatory environment. I do not think we want to go down that route when it comes to AI as well as other areas.

Stephen Metcalfe: And that would be your recommendation or point to the Government.

Hugh Milward: Yes.

Stephen Metcalfe: Jen?

Jen Gennai: For international assistance, I would look to the bodies I mentioned earlier, like ISO, IEEE and the OECD, which have already brought in so many diverse perspectives on learning about what the common areas of concern are, such as fairness and accountability, instead of trying to start from scratch and establish the UK's own, while still allowing, as the UK has already started to do, and I think we are aligned on this, a principles-based approach that allows for innovation while managing the risks.

Adrian Joseph: I would say Singapore, and some elements of the EU are worth looking at—I would not rule them all out in a blanket way. The one thing I would ask the Government is to allow British companies and sovereign capability to develop, because if we do not get behind generative AI and large language models, the big tech companies have cloud capability and the expertise, and we will have handed over control in a very significant way. That would be my biggest request and suggestion to the Government.

Stephen Metcalfe: Perfect. Thank you.

Chair: I thank all three of our witnesses for an excellent first session today. It has overrun considerably but that is simply because of the interest in the answers and the different paths we have had to go down in terms of innovation and the related ethics. Thank you all very much.

Examination of witnesses

Witnesses: Professor Dame Wendy Hall and Professor Sir Nigel Shadbolt.

Q144 **Chair:** We are now joined by a panel of academics. Joining us virtually we have Professor Dame Wendy Hall, regius professor of computer science, associate vice-president for international engagement and an executive director of the Web Science Institute at the University of Southampton. She is also a fellow of the Royal Society, she was co-chair of the UK Government's AI review, published in 2017, and she serves on the UK Government AI council, together with Adrian, whom we just heard from.

Joining us in the room we have Professor Sir Nigel Shadbolt, professorial research fellow in computer science and the principal at Jesus College at the University of Oxford. Professor Shadbolt has been researching AI since the late 1970s, so his thoughts on the curve will be interesting. He is chair, and co-founder, with Sir Tim Berners-Lee, of the Open Data Institute, which works with companies and Governments to build an open and trustworthy data ecosystem.

We are very glad to have you, Professor Hall. Sorry for the delay, but I hope you were listening to the first session, because it was fascinating. Together with Jerome Pesenti, the CEO of AI drug discovery company BenevolentAI, you wrote the independent review of how to grow the UK's AI industry. Graham Stringer is going to ask you specifically about your recommendations, but given that that was 2017, I just wondered

whether we are collectively where you thought we would roughly be in 2023, in terms of the speed at which AI has developed and whether we are in the right place in the UK at the moment.

Professor Hall: In terms of where the speed of AI has developed, as the previous panel said, it is hard to know where we are on the curve, but I am very pleased at how the UK has taken forward the review. As of a result of that, we became a sector deal, the AI Council was set up, the Office for AI was set up, we had a new road map in 2020, and in '21 we had a new strategy launched, which we are working through now.

The key thing I want to say is that we need to keep that going. We are still at the beginning. Even ChatGPT is not the end of the beginning; we are at the beginning of the beginning with AI, even after all the years it's been around, and we need to keep our foot on the accelerator and keep ahead of the curve as much as we can. Now is not the time to stop.

Q145 **Chair:** Perhaps I could ask you both for your reflections on the first panel—I saw that Professor Shadbolt was in the room. The big tech companies that we heard from are obviously full of innovation, but do you think they are taking the risks of AI seriously enough? Perhaps I can come to you first, Dame Wendy, and then to you, Sir Nigel.

Professor Hall: There are two things I wanted to say. It was a very interesting panel. I think that, BT being very based in the UK, we need to take very carefully what Adrian was saying; he had some very wise words for you.

We are not going to compete with companies like Google and Microsoft now in terms of what they are doing, but we absolutely need to develop a sovereign large language model capability. There is a meeting at the Turing today with 300 people attending and we have pitched to ARIA, but it really needs the UK Government to get behind this; not in terms of money—the money is out there—but we need the Government to get behind it, because otherwise we will be in the same place with this as we are with the cloud.

What didn't get mentioned on the panel was that what we see with ChatGPT—all the biases and things it gets wrong—is all based on searching across the whole internet. When you think about taking something like a generative AI system technology and applying it to NHS data—data that, as far as we know, is correct; we can trust it; we know what we're doing when we're training it—you can see that that is going to be incredibly powerful. Do we really want to be reliant on technology from outside the UK for dealing with something that can be so powerful in enabling the UK to access all that health data, both for the UK and for the rest of the world? It is the same with our intelligence services; they are getting very behind this as well. So that is something: the scope of how you train the algorithm has to be taken into account, and I think we could do that really well, and we need to do it well.

The other thing they didn't mention is that China is—I don't know if it is way ahead, but China has its own large language models. Baidu is doing

one, Alibaba is doing one—I think I've got the right companies—and the Chinese Government itself is doing one. I believe that one's called Wu Dao, and the last I heard—it's hard to get the information out; I haven't been to China for a while—that was more powerful than GPT. So don't think China is not doing its own thing here: it is.

Q146 **Chair:** Thank you, Dame Wendy. Professor Sir Nigel, what are your reflections on the first panel and whether big tech is being a little bit blasé or not?

Professor Shadbolt: I thought it was very informed, and it was encouraging to hear that they are alert to the issues that are going arise here. I would just say, in the broader context—having gone up to Edinburgh in 1978, to the only department of AI in Europe at that stage—that AI has been around for many decades. In fact, the first great flourishing was the Alvey programme, funded in the UK in the 1980s in response to the Japanese fifth-generation initiative, when everyone was terribly worried that the parallel computers that they were looking to develop would put manufacturing in Japan well ahead of everyone else in the UK, the US and Europe—to some extent, all responded.

We have been through those various curves in the past. AI is on a continuous trajectory, hugely due to the fact that with the exponents of change—particularly in data, if it is available, and I would want to come back to that and to talk about the size of the prize, the data that is available and what we need to do about that—we have the materials engineering. That was well reflected in the comments of a number of panellists. The extraordinary increase in computer power and processing power has enabled, with data availability, these large language models to come into existence. We kind of knew that was a possibility decades ago, in some sense, but the inevitability of those exponents, which we can see going out well beyond where we are now, is the thing that should give us cause to think hard about regulation and what we need to do to put the guard rails in place.

I thought that was informed, and I absolutely echo what Wendy was saying about the capabilities that we need. AI has flourished because of lots of independent research in great universities and research institutions around the world. Back in the day of Alvey, we had our own sovereign capability—companies that made computers, like Marconi or ICL—but we have to remind ourselves where we are now and what is needed to help substantiate an ability to compete both in hardware and in software terms.

That requires us to think about investing in sovereign capabilities. One group at risk of being left behind is university researchers. That is where much of this research was done originally, because their access to the kinds of compute you need is not organised terribly systematically. In the past, when we have thought about that, in high-performance computing, where the UK has tremendous capability—its ability to do scientific modelling is premised on that—we have had to ask ourselves what we do and can procure. Some of the companies that we know of will give us

access to that material, but we have to think about how we can sustainably guarantee our access to that compute power.

Chair: Thank you, Sir Nigel. I am pleased to say that our Chair, Greg Clark, has joined us fresh from his legislative exertions. We will come to Greg in a little, but we will go first to Graham Stringer and then Stephen Metcalfe. I would like to wrap this up by about quarter to 12, if everyone can bear that in mind.

Q147 **Graham Stringer:** Professor Hall, if you walk around this building or, I dare say, every university, office and factory in the country, you will find them awash with terminals and computer power. If you look at the economic growth in this country since that huge increase in computer power, and the uses of it in industry, academia, politics and everywhere, you find it very hard to find any impact on economic growth. Yet, over the next 13 or 14 years, because of AI, you have been predicting growth of considerably more than half a trillion pounds. Does that mean AI is different, or will it do things that just computing power did not?

Professor Hall: That is quite a hard question. I didn't—

Graham Stringer: It is an important question.

Professor Hall: I think it is an important question too. In the world of digital, the UK is a huge leader. Our problem, as we have talked about, is sovereign capability. Consider that if we don't do it, we will just become a service industry country.

When we did the AI review in 2017, we were charged deliberately with driving economic growth, because at the time the rhetoric was all about AI taking everyone's jobs. In some countries, that is a huge concern, as more and more automation comes in and people in rural areas cannot get other jobs. In the UK, we can; we can harness the technology and we can use it to drive the economy and to grow jobs. We have started it, and we have to keep our foot on the accelerator to do that.

The skills piece that we launched after 2017, which the Office for AI implemented, includes considerable investment in our universities and MSc courses to create the skills we need to harness that. As Adrian kept saying, we need to make sure people have the skills to work in this industry. We need diversity of skills because there will be lots of new jobs—for example, AI auditors: people who will look at and manage regulation in the AI industry.

The opportunities offered by AI are huge. I am not an economist, but I think there are two reasons why, as Adrian said, we are third in the Tortoise index for AI—third to the US and China. First, we have the university legacy, as Nigel said. The UK has fantastic researchers in some of the top universities in the world, particularly in AI; we have been leading in this for a long time.

Secondly, we have a fabulous start-up culture. We have twice as many technology start-ups as the rest of Europe put together, and a lot of those

start-ups these days are AI start-ups. But our biggest problem is that we do not scale up. We get a lot of unicorns, but we do not scale up. To me that is one of the most important things I can say this morning. The Treasury needs to help create incentives for the investment to go into the unicorns to create big companies here. We can do that if we create the right incentive environment. At the moment, our companies have to go to the US to get investment for scale-up. I am not saying we will ever create a Google or a Microsoft, but for sure we need to create growth in this sector. I do not know whether that helps at all.

Q148 **Graham Stringer:** Interesting. Is there any particular part of the economy you think AI would give a dramatic boost to? You made recommendations in four particular areas in your report to the Government. Have the Government implemented those recommendations? Are you pleased with their response?

Professor Hall: Yes, in the sense that the Office for AI has implemented what we recommended, but it is all about take-up in industry too. It is about creating the climate to enable industry to use AI. I always answer this question by mentioning health, because we all know the problems every country has with running a health service. It is huge on our agenda. I think AI could make huge savings in the health service, and could free up the doctors and nurses—the frontline people—to do what they need to do in terms of the patients, while the AI does the analysis and the data to create the efficiencies in running the service. We could also do a lot with the NHS data if we could get that right, in terms of creating wealth for the nation.

Q149 **Graham Stringer:** Thank you. Professor Shadbolt, I saw you looking quizzically at my first question. Do you want to respond to that question about the economy and why we have not seen the benefit of computing power?

Professor Shadbolt: It is quite a contested area—this notion of "How far has IT and computing in general accelerated and enabled economic growth?" It would be odd to imagine that it has not had a huge impact on jobs created, new opportunities, content created and new insights. It can sometimes be hard for the metrics we use to estimate those things in terms of pure growth and to capture absolute insights, but I cannot imagine any area of our life where value has not been hugely added by the pure dint of compute power. I give you the example of the use of genomics, and the analysis of a whole slew of our health and biological data, off the back of computing; that has created new classes of therapy, new industries and new businesses.

Q150 **Graham Stringer:** Do you think that as we move to using AI more, as it is developed, that is increasing or decreasing transparency?

Professor Shadbolt: On the principle of transparency, one of our motivating principles at the Open Data Institute is accountability and transparency. In fact, one of the key messages I would like to impart is that although we are talking a lot about AI, for the algorithms their feedstock—their absolute requirement—is data. The need to build a

trustworthy data ecosystem is paramount. We have had a long leading tradition in this country of endeavouring to make public sector data as open as possible. That is a foundation on which you can build lots of services, and many of them are going to be driven by AI.

As we heard in the previous session, with the advent of these new large language models some data will need to be very highly curated and managed. We might want to come back to the health data, because I think that there are some interesting innovations in health data in the UK right now, some learnt from the pandemic. We would not want to lose those opportunities for how we can do AI-accelerated analysis and drug discovery off the back of that.

Coming back the point about ensuring that our data is high quality and that we have a good data ecosystem, we are open as a foundation for as much of our data ecosystem as possible. That helps with transparency and accountability, of course. These new language models are ingesting huge amounts of data from the web, from Wikipedia, from books and from scientific literature and elsewhere. They know that there are real challenges around the inherent biases that were talked about earlier and the structure of that data. As for the next generation models that will be launched, which are in development right now, one of them, the PaLM model, has 50% of its input from social media, just because there is so much of it in so many languages to train the model. You can immediately imagine what we have got to worry about with that data landscape.

We need data that is accessible—that will help for transparency—available and assured. Where it needs to be protected, we have to make sure it is protected, and that brings us on to how much of what you will be thinking about with AI regulation and governance will also need to dovetail into and relate to the work being done on the data Bills that are being developed and are under review in Parliament.

Q151 **Graham Stringer:** I think I am being fair to the previous panel in saying that they said that you cannot completely take bias out of the data, but you can gradually improve it. Do you accept that?

Professor Shadbolt: In some real, fundamental sense, all data is collected with a purpose in mind. That purpose might be entirely obvious, it might have built-in limitations about how it can sample and what it can sample. There was a rather nice description from the previous panel of how the bias can arise in the data itself, at the point of implementation and from the people who design the actual algorithms. We should pay attention to that notion of where bias sits and what we mean by fair in this context; there is not some kind of simple, out-of-the-box description. It is about in what context, for whose purposes.

This whole area is an enormous area of research. We also often hear an appeal to the fact that these will be black boxes, non-explicable. I think that that is a counsel of despair. We absolutely have to demand that these systems begin to render some of the processes by which they came up with the output they do more transparent and more explicable. There is an

active area of research in machine learning and deep neural networks that is seeking to build explainable AI systems, and what we should not do is imagine, "We've got what we've got, we will have to put up with it."

Q152 **Graham Stringer:** That takes me on to my final question. We are told by some of the evidence that the people who have written these algorithms do not quite understand how the machine has got to where it has got to, which is, I think, the point you are making. Two questions follow from that, I suppose. Are you saying that that is just where we are now and that we don't have to accept that—that we can understand everything that is going on in the machines? Secondly, on where we are now, what are the implications for scientific research, when the nature of doing science means you have to be able to check what has gone on, if you do not really understand what the machine has been doing?

Professor Shadbolt: That is a really good question. We worry about ChatGPT being used in various contexts, but one about which there has been quite a lot of thought and examination is writing scientific articles and doing scientific research. On generative AI, the types of models that are used to explore science are going to be an absolute feature. Again, this is a rather nice description: we are entangled with our machines now. They are augmenting our intelligence. We need to celebrate that and work out how that can be done in a way that is safe and scrutable.

I should say that I am absolutely dyed in the wool and committed to the scientific method. We know that there is a bit of a reproducibility crisis in science. It is not always the case that you can reproduce straight out of the box the results in other papers in the way that you might have thought. We should place demands on ourselves to try to do better in this new booming area of AI—AI applied to areas like scientific discovery. It offers opportunities, and we should ask ourselves how we can push that forward.

You talk about understanding every aspect of what is going on, but in some sense these large models have a fundamentally statistical character, so there will always be some sense of indeterminacy in them. We need to look at the overall ways in which they come to their conclusions and where their frailties are. That is why this sovereign notion is such an important one. We have to be able to kick the tyres on these systems just as hard in our university research groups as the industrial researchers do in the big tech platforms.

Of course, we want to enjoy and look at their research, but to really do this, we should do the same thing that we find in safety-critical areas such as civil aviation. Avionic software is put through extraordinary testing regimes to ensure that it performs in an expected way. We won't hold all our systems to that level of accountability, but we should think about how we institutionalise that.

Professor Hall: Just quickly, I agree with everything that Nigel said. Adrian used a wonderful phrase: the ultimate black box is the brain.

Human beings can't always explain how they have made a decision and why they have said something.

There are three entities involved here. We need more research on explainable AI. Companies like the Googles and Microsofts need to be part of that, and we need to think about how we develop systems that can explain what they do. The Government has to play a part, in terms of regulation. I like to talk about auditing. We can't wait for the research to be done to regulate AI. I like to think about the whole idea of how we audit systems that you can't explain. You can test them and simulate them in various environments.

I would also like to bring people into this. Citizens' juries and citizens' groups would make it easy for people to raise issues and report bias and problems without being pilloried in the media for being whistleblowers. We should get away from that language and make sure we involve people so that they can report things easily and transparently.

Chair: Thank you, Dame Wendy. On Sir Nigel's point about the reproducibility crisis, you may both be aware that this Committee has an open inquiry on that. We looked at AI as part of that, and we should be reporting on that soon. Katherine Fletcher has a very quick supplementary, and then I will bring in Stephen Metcalfe.

Q153 **Katherine Fletcher:** Thank you both for coming. It is appreciated. I am listening to you on testing regulation. If I may be slightly cheeky, it is almost like we are trying to regulate AI in the way we always have done—almost an analogue way—which is build something, test it and make sure it does not break. Is there a role for turning that on its head and for regulators to say, "This is what this thing has to adhere to as it is working itself out"? In my simplistic world, I would call that input rules: it has to do a read-out here, it has to understand that it is powered by electricity, or whatever. Could we get a global standard of input rules for AI, or am I missing the point?

Professor Hall: I think that's great. I should have said earlier that the Office for AI will be producing the White Paper on AI regulation in the next month, and you will all have a chance to scrutinise that. I don't believe they prescribe in that exactly the answer to your question, but on the idea that we develop ways of doing that, I think the UK could be a real leader in doing this in an input/output way.

Q154 **Katherine Fletcher:** It is almost the AI ten commandments: thou shalt not kill, thou shalt give a read-out of what goes on in the black box.

Professor Shadbolt: I think that is where some of the ethics in AI work is trying to make progress. It is a little bit like what medical ethics did as it came into view. It developed concepts that didn't really exist, such as futility of care, or the notion of brain death or of informed consent, which actually originated right back in those early deliberations.

Katherine Fletcher: Or the notion of, "You can't steal electricity to run your servers off a vital human system."

Professor Shadbolt: Exactly. I think thinking in that way would be helpful. We have heard a lot about looking for the harms, rather than trying to be too normative and imagining ahead of time all the things that AI could be put to use for. Clearly, AI is a general-purpose technology. It can be used for ill or good; it is dual use. All that being said, our regulators—many of which are trying to gear up for this world of AI, from the CMA to Ofcom and so on—are trying to understand what technical skills they need and what they would need to put into the process to enable their job to be easier.

Again, I come back to the challenge of AI in the mobile data ecosystem. All the data flowing off your phones at the moment is being used in all sorts of ways to model behaviour to understand your preferences. It is very hard to know where that data goes and where it flows, because there is an arms race to encrypt the apps that do that work. There might well be an opportunity for a piece of regulatory technology that says, "Thou shalt not encrypt the destination of the packets unless there is an overriding security reason to do so." It is a good thought.

Q155 **Katherine Fletcher:** I am very conscious of the time. If you would like to give us the fourteen commandments for AI, perhaps by writing to the Committee, I would be interested to hear what those input conditions should be.

Professor Hall: You know the House of Lords has the ten commandments, which were developed by the bishop who was on that—

Professor Shadbolt: I think they're very helpful, but there are some particularisations that could be updated.

Katherine Fletcher: I'd like to get to an implementable framework that you could choose to make legal or not.

Q156 **Stephen Metcalfe:** Good morning, both. Sir Nigel, I think you recently wrote that data "should not be monopolised and controlled by those who collect it, but placed in the public domain for wider societal and economic benefits." How do you think that is going? Do you think that particularly big tech companies who collect huge amounts of data, as you have just mentioned, through devices and so on are as committed to that as perhaps public bodies are trying to be?

Professor Shadbolt: Thank you for that. I am not an open data revolutionary in the sense that all data has to be made openly available. Clearly there is a whole spectrum of data from personal to sensitive to shared through to what could be open. The general presumption is that a huge amount of data that is collected could and should be made open for wider benefit. It is often forgotten that Google's original world-beating search technology was born of the fact that they harvested the webpages and in the webpages were all the links and those links were put there by human beings. All that human ingenuity that was out there in the open allowed them to build a great search engine.

There is some kind of notion that as data is collected, can't we ask ourselves whether it might be much more valuable to have a public as well as a private good orientation? How's that going? Mixed.

Q157 **Stephen Metcalfe:** Mixed; okay. Does it require intervention, therefore, if it is mixed? Or will it come about?

Professor Shadbolt: I think, again, you can see that it is about aligning incentives. We saw a great example of where making it open really worked, and this was during the pandemic. Scientific publications were made available for data mining. That was not allowed previously. The insights that flowed from that were really great. Scientific research and data is a great example—it has been for a while—of where the sharing is multiplicative; it really does work in that kind of way. However, we have to understand that sometimes research data needs to be held close, needs to be kept private and needs to be held in a trusted research environment. Here we look at opportunities to do that, which we have learned from the pandemic.

There is an extremely interesting project called OpenSAFELY that is run by Ben Goldacre, who did the Goldacre report for the Department of Health and the Secretary of State. There, the really sensitive NHS patient data that Wendy was talking about stays behind the firewalls and is linked, but your analysis goes behind the firewalls, looks for patterns, brings back results—again, anonymised at an appropriate level—and it stays secure. That is an extraordinary facility to maintain and exploit. Powered with AI, it could be enormously beneficial.

Q158 **Stephen Metcalfe:** Thank you for that. I think you touched on what the risks of having too much open data might be. To ensure that those risks are not given more credibility than they need, do we need wider public understanding of what we mean when we say "data"? Actually, data is a whole range of different things.

Professor Shadbolt: That is the fundamental challenge of this whole programme of work. Wendy's report talks about the data and literacy piece. Where do we need to inform ourselves—at every level of society, and into our schools—about the new reality, what data is, what its inherent properties are, how it can work, the various prescriptions you need to put around it, the sensitivities and the insights you can get? Again, the Open Data Institute has been developing what we call the Data Ethics Canvas, which takes you on that journey. That journey is something that I think all of us have to have a sense of. We will not all be data scientists on this stuff, but we should have some appreciation of how that landscape can work and operate effectively.

Q159 **Stephen Metcalfe:** I just want to touch on the issue around accountability, transparency and repeatability. To get to a point where algorithms are more accessible, transparent and predictable, do you think that would lead to a compromise of a developer's IP, or can we completely separate those? You do not have to share your trade secrets to be able to be transparent.

Professor Shadbolt: I think the fascinating thing about this area is just how much of what has driven development is actually open-source code in any event. In fact, one of the most successful companies in this space is Hugging Face. Its entire business model is to collect these models together, make them available and have innovation happen around them. I think that very often the secret sauce—there is of course some—has as much to do with the hardware configurations that you've got as the algorithms, famously, and the data you have available.

We can and should think about this notion of what is made available—I agree—and there is a whole range of responses we could have to that. Rendering it transparent does not mean that you have to throw your secret IP components over the fence. It may be that you will be asked to give a functional characterisation of how the algorithm is working, without revealing how you have perhaps optimised it in some way. It might be that you could look at having a set of benchmarks against which you are going to require the system to perform.

On this idea of explainability, when we ask the experts who inform us for explanations, the types of explanation they give are not one kind of thing. There is a whole body of literature around what effective explanations might be or could be, and we need to think more widely about how those are engineered into our interactions with our machines.

Professor Hall: I just wanted to say a couple of things. One is that we have not mentioned today the fact that the ChatGPT algorithms are being trained on the entire internet and are compromising a lot of people's intellectual property. We have laws to deal with this. There is litigation already started in the States by the artists, the poets, the novelists and the musicians whose IP is being compromised and used in ChatGPT in completely anarchic ways. We have laws, we are ready to deal with this, and we need to make sure they are applied in this area. I am sure that companies are thinking about this a lot, but that is going to be a big issue going forward.

The other point that I want to make—I think you got this from Nigel—is that the Data Protection and Digital Information Bill is just as important for AI as the AI regulation paper that is going to come out. You cannot disconnect those two things. The regulation around data will affect the regulation in AI.

Q160 **Chair:** To follow up on what you just said, Dame Wendy, about the creative industries in particular—obviously, it also affects law and all sorts of other areas—are the laws that we have sufficient? There has always been protection for pastiche and parody. Is it not possible that people will just argue that that is what ChatGPT is doing?

Professor Hall: I am not a lawyer, but we went through it with the way we get hold of music on the internet. We will have to evolve this as the AI evolves. I am sure that if you were an artist and you saw your painting being used in a ChatGPT answer, you would want to have some recompense.

Q161 **Chair:** Yes. I think the issue might be the difference between your painting and something in the style of your painting.

Professor Hall: I think that is where the lawyers are arguing at the moment. Musicians will be a big part of this.

Chair: Thank you, that is very helpful and interesting.

Q162 **Dawn Butler:** Wendy, I will start with you. I have a couple of quick questions. Are there any aspects of how the human-smart machine relationship works that cannot be addressed by regulation? How might the Government deal with that?

Professor Hall: The diversity angle is really worrying to me. It has been throughout my career in computing, but it is more important for AI because it is going to profoundly affect society, which is very diverse and needs to be diverse. We need diversity in the people who are working in AI and designing the algorithms, and we are not getting enough of that at the moment.

We also need diversity awareness in the people who are applying to be regulators, thinking about the risk, helping businesses adopt AI and helping the NHS use AI. One of the wonderful things that came out of the AI review was the funding of scholarships for MSc conversion courses, which allows people from a non-STEM background—particularly from under-represented groups in the technology field—to get scholarships to study new MScs that the universities are creating. If anything, that is the thing I am proudest of. We need more and more of that as we move forward.

Q163 **Dawn Butler:** Thank you. How concerned are you about AI amplifying the biases that already exist in the NHS?

Professor Hall: That is what I said earlier. Yes, I am very concerned. As you heard from the previous panel, it is really difficult to get all that out of the system, because it is there and you cannot retrospectively do the clean-up. Going forward, I would really involve citizens groups in this. We asked for that in the review, and a lot of the reports are saying that, as we start to adopt AI in business, and particularly in parts of the public sector like health, we absolutely need to include citizens groups and say, "This is what we are going to do. This is how we are going to do it. This is how you can report something that you feel is unjust or biased in a way that is easy, open, transparent and makes people feel included."

Q164 **Dawn Butler:** Do we need to look at the bias that already exists and accept that it exists before we put it in the system? The body mass index is an easy example that I always use, because it has an inherent bias: it is based on men, not on women or any other groups.

Professor Hall: As I said, you cannot retrospectively go through and clean up the whole system. You would wait forever to do that. It has to be evolutionary, so that we recognise as we go forward where the biases are coming in. Like the companies were saying, you go through and fix it

when you can. If we wait forever to clean everything up, we will not ever start.

Q165 **Dawn Butler:** I find that problematic. The health outcomes, because of the bias that exists in the system, ultimately lead to a disproportionate amount of deaths, for example. I think we must ensure we clean up the system as it exists, and that is paramount to what we do.

Professor Hall: We need a process for doing that. I would always argue that the human has to be in the loop, so that any decision-making process has to include humans in the loop.

Q166 **Dawn Butler:** Thank you. I have one more question for you, Wendy. Is this statement true or false: you can manipulate generative AI?

Professor Hall: Gosh. I am an academic; I cannot give—yes, you can.

Q167 **Dawn Butler:** Thank you very much.

Professor Hall: I always say, "What does Nigel say?"

Q168 **Dawn Butler:** We are going to Nigel next. Nigel, true or false?

Professor Shadbolt: There is a whole field called prompt engineering where you seek to get it to give better answers or different kinds of answers. Of course, it is how you tune and adjust the model. ChatGPT is the thing you get in front of you, but there are lots of ways of getting into and behind those models and changing their behaviour, for sure. There is not a canonical set of outputs from these things.

Q169 **Dawn Butler:** So it is true.

Professor Shadbolt: With all those caveats. To your earlier point, you are absolutely right. That wonderful book by Caroline Criado-Perez, "Invisible Women", is all about this fact that data is collected that already inherently skews all sorts of outcomes, whether it is road safety data or mannequins in cars and all those things. If nothing else, the industrialisation of AI in this context is lending new urgency to those questions.

Q170 **Dawn Butler:** Absolutely, and if we do not clean it up, we are just embedding it in the system.

My last question to you, Nigel, is do we look through rose-tinted glasses and hold the UK on this pedestal with regard to, for instance, democracy? We are ranked 18th in the democracy index; with regard to a free press, we are ranked 24th. If we say that whatever we create is superior to other countries, is that through rose-tinted glasses?

Professor Shadbolt: It always helps to be slightly humble in what you are claiming, because imagining that other people might be doing it better is a good heuristic, or imagining that you are not doing everything as well as you might.

I think the proposition you want to drive with the things you put in place, your public services and the way these machines evolve their algorithms

to be at our service and support human flourishing in an age of AI—that is a phrase I like to use, "human flourishing"—is about values. It is about trying to be quite careful to keep revisiting to what extent those values of transparency, equity and access are being maintained. That we recognise.

One thing we have not talked about much today—and I am sure you want a bit of context—is the ability for these adversarial AI systems to fool one another, to generate deepfakes and to actually subvert the very processes we hold dear. There is that arms race. There was a very interesting question from Katherine earlier about the idea of self-serving systems. AI goes all the way down the species tree in a sense; at its simplest level, some of modern malware—the kind of things that infect and replicate themselves across systems—is driven by machine learning insights and insights from AI as well.

There is an absolute game-on here for us to pay attention to where we are and how we do what we do based on values.

Dawn Butler: Thank you.

Chair: Thank you; final questions from Greg Clark.

Q171 **Greg Clark:** Thank you, Chair. Thank you to Sir Nigel and Dame Wendy for giving evidence. I am sorry to be late to my own Committee, but I can tell it has been a fascinating session.

Just thinking of the Government's regulation of AI—perhaps I will come to you first, Dame Wendy—is there an opportunity for the UK to innovate with a particular approach to regulation, perhaps frontrunning some others, or should we be aiming at a more consistent international approach?

Professor Hall: No, I think there is a chance for us to be world leaders here. The eyes of the world are on us and the regulation paper that is due out next month. I have not seen it, but there was the policy paper last July that was very well received, and the AI Council has been asked to advise on particular points in the regulation paper.

It is a different approach. It is principles-based and focused on the context in which the AI is deployed, enabling innovation. While in the end there is a risk-based element, it is a very different approach to the EU AI Act. Adrian mentioned Singapore earlier; they are looking at this, too. But a very small number of countries are really getting their act together in terms of AI regulation. As I said, I think the eyes of the world are on us. We do this stuff really well. We have done it in the finance industry. I think we can do this. It is one of the things that keeps us up there, third in the table. This is something we can really lead the world on.

Greg Clark: Thank you. Sir Nigel?

Professor Shadbolt: I think there is a lot to that—absolutely. That ability to use our whole legal tradition of taking a harms-based approach rather than trying to anticipate with these omnibus kind of regulations is a real

feature in an area where you do not quite understand the unintended consequences of deploying these new technologies.

One point that was made in the earlier session that I would support was that we do have to pay attention to what is going on around the world. Wendy has mentioned Singapore. Estonia has interesting work, perhaps unsurprisingly in this area given their other tradition of work in digital government. The standards piece is super-important. We have established an AI standard, but I think we need a data standard so that these two work cheek by jowl. We need to be in these bodies—OECD, ISO, IEEE, W3C—where those questions are being framed and addressed. It is often unglamorous work being on standards bodies, but it is absolutely essential.

Q172 **Greg Clark:** In the spirit of humility that you recommended in your answer to Dawn's question, are there other countries, either through those international bodies or unilaterally, setting themselves up in a way that we can learn from?

Professor Shadbolt: It is noteworthy that China did produce its own AI regulations. They were algorithmically focused and different in style. They have also invested a huge amount of effort in standards work now. That calls to attention the fact that we probably want to do that, too. There is some very interesting work as well linking back to how we procure AI systems in Government. On that procurement piece, again I think there are some very interesting insights that we can get from other countries, and from our own practice. Back in the day we were committed to open contracting. There is a whole set of interesting capacity that we have in the UK to bring transparency to the process by which we want to procure systems that we hope will be transparent and adhere to the kinds of regulations we will see emerging with our Bill. On marrying these elements together, the challenge for us is to ensure that the sum is not less than its parts.

Q173 **Greg Clark:** Finally, in summary, our inquiry makes recommendations to the Government on the governance of AI. What would be your top recommendation to the Government, through the Committee, when it comes to designing and optimising the future governance of AI?

Professor Shadbolt: We have majored on supporting data literacy. I think that foundation is super-important. Do not forget the management and organisation of the data ecosystem that will drive all of the AI applications in the future.

Greg Clark: Thank you. Same question to Dame Wendy. What would be your top recommendation?

Professor Hall: I would say we have to be more ambitious rather than less. There is a sense of "Well, we've done AI. We had a review, we had a strategy, and we've got a regulation paper." This is just the beginning. As I think I said before you arrived, this is the beginning of the beginning. We need to build on what we have done and make it better and world leading, and we need to think about sovereign capability.

Greg Clark: Thank you.

Chair: May I thank both our professors for their evidence in this session? Together with the first panel, we have had, over two hours, fascinating evidence. We are all looking forward to the regulation and White Paper that you referenced a number of times, Dame Wendy. And we are all looking forward to the rest of the inquiry. We have a lot of work to do. We have had a huge amount of food for thought today. I thank both of you and the three earlier witnesses.