HOUSE OF LORDS

Communications and Digital Committee

Corrected oral evidence: Large language models

Tuesday 19 September 2023

2.35 pm

Watch the meeting

Members present: Baroness Stowell of Beeston (The Chair); Baroness Featherstone; Lord Griffiths of Burry Port; Baroness Harding of Winscombe; Baroness Healy of Primrose Hill; Lord Kamall; Lord Bishop of Leeds; Lord Lipsey; Lord Young of Norwood Green.

Evidence Session No. 2 Heard in Public Questions 12 - 20

Witnesses

I: Dr Peter Waggett, UK Director of Research, IBM; Dr Zoë Webster, Director of Data and AI Solutions, BT; Francesco Marconi, Co-Founder, Applied XL; Dr Nathan Benaich, Founder, Air Street Capital.

USE OF THE TRANSCRIPT

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

Examination of witnesses

Dr Peter Waggett, Dr Zoë Webster, Francesco Marconi and Dr Nathan Benaich.

Q12 **The Chair:** This is the Communications and Digital Committee, meeting again as part of our inquiry into large language models. We have two panels of witnesses this afternoon. We will try to divide our two sessions, the first panel covering what I might loosely describe as opportunities, while the second will be predominantly about risks. Clearly, we will get to fundamental risks with the second panel, but in this first session we are keen to hear about the benefits of large language models. Where we do discuss risks—I am sure they will come up, because it is hard to talk about benefits without addressing risks—then we will be very interested to hear from our witnesses today about how these risks are being, or can be, mitigated to give people confidence in large language models.

I should add that we do not necessarily want all four witnesses to answer every question that we ask because we have quite a lot to get through. Before I get to the questions, let me invite each of you to introduce yourself and the organisation that you represent, if you are representing one. In doing so, perhaps you can use one word to define yourselves, as in whether your business or organisation is a deployer or an investor. That is just to give anybody watching this a sense of where you are all coming from.

Dr Zoë Webster: Thank you very much. I am AI director at BT, which is a big telecoms company moving into more technology. We provide a lot of the critical national infrastructure of the UK.

The Chair: You are a deployer?

Dr Zoë Webster: Yes, we are a deployer.

Dr Peter Waggett: Thanks for the opportunity to talk to you. I just want to go through the two roles I have that I think are relevant to what you are discussing today. First, I am research director for IBM in the UK and I am working with teams of IBM employees that are delivering one of only three global discovery accelerators that we put in place.

It is also important to say that I am on the executive board for a project called the Hartree National Centre for Digital Innovation. This is a jointly funded project between IBM and the UK Government to deliver AI and emerging technology skills and facilities to UK companies. From my perspective, I believe that this is a really good example of the public and private sector working together to develop AI.

The Chair: IBM is a deployer, yes?

Dr Peter Waggett: A deployer and developer.

Dr Nathan Benaich: It is a pleasure to be here. I am a venture capital investor. I focus on early-stage technology companies in machine learning and AI technology across tech and life sciences. I have been doing this for 10 years or so. I did my grad school in the UK, mostly in computational biology in cancer research. I also produce the annual

State of AI Report, which is our open-source document that tries to appraise progress in research and distributing technology, geopolitics and other sectors.

Francesco Marconi: Good afternoon. I am the chief executive at Applied XL, which is an AI in computational journalism company based in New York City. We are a deployer of artificial intelligence in addressing issues related to accuracy, as in implementing editorial processes for vetting the outputs of AI. My background is in computational journalism, previously holding positions at the *Wall Street Journal* and the Associated Press.

Q13 **The Chair:** Thank you all very much for being here and giving us your time. Perhaps I can start by coming to you, Dr Benaich. Could you tell us what you see as the main opportunities in the UK for large language models and perhaps what you see the investment and adoption landscape looking like over here? Just give us the headlines, as it were.

Dr Nathan Benaich: From a headline perspective, AI has probably been in development in this country for decades. Many of the leading lights in this space who are either still here or have emigrated elsewhere trained and learned their craft in the UK, so it continues to be a hotbed of intellectual spirit and development.

Practically speaking, a lot of the applications tend to focus on the areas that the country has historically been very good at: areas such as cybersecurity, life sciences, healthcare, enterprise software and fintech. In general, on a global perspective, the US dominates with regards to company formation, investments, acquisitions and so on. The UK tends to rank third, behind China. To give you some stats on this, if we look at generative AI in this topic of LLMs we are discussing, in the last four years or so we have seen $6 billion invested in just San Francisco companies alone. By contrast, in London we see around $365 million. That is about a 20-fold difference.

It is the same story for AI chips—I think we will discuss some of this today as well—where there is roughly a 20-fold difference between investment in China, which is around $7.5 billion in the last four years, and $450 million in Europe entirely. There is a lot to play for, but at the same time the centre of gravity is increasingly shifting west, so we have a lot of opportunity to build up the domestic ecosystem here.

The Chair: On the investment stats that you are giving us, are they investment in the development of AI or in businesses that could deploy the technology?

Dr Nathan Benaich: These figures are in companies that are developing either products powered by machine learning or machine learning-based services that other companies can use to create their own products in house. It is anything from a large language model provider to a developer tool that helps indie developers to get AI apps off the ground, or a company that has machine learning baked into its product.

The Chair: Dr Waggett, do you want to give us your perspective on the

landscape?

Dr Peter Waggett: The first thing to say is that we are working with companies in the UK now through the Hartree centre on large language models. One of the projects that we have done as a team has been based around Wimbledon. Everybody is familiar with Wimbledon getting statistics about speed of serve and all these different things. We did an interesting project where we basically took written summaries of matches against an agenda of, "What makes great?". In that instance it is looking beneath the statistics to understand what people described as the attributes—the things about role models that made them important tennis players. We are doing work in that space. We are also doing work with a lot of retailers, using effectively large language models to help them with aspects of their customer service.

The more important thing, which is behind the question, is what will happen in the next couple of years. Again, I would come back to the Hartree centre, which we are involved with, where we are starting to get involvement from large companies trying to understand how they can make sense of this. In particular, with our agenda of accelerated discovery, we are looking at how we can speed up the time to market for new materials or new drugs, but also try to reduce the cost for them.

One aspect that we are playing into with the large language models is being able to analyse scientific literature. On that basis, we can give researchers an insight into what is working in their field and what is not working, taking account of the fact that the AI can help. All researchers will probably admit that they just cannot keep pace with the type and the quantity of material that is being produced, so we think the accelerated discovery aspect of work will be very big in the next few years. We can work with companies in the UK through the Hartree centre and help them do that.

The Chair: We will come on to applications in a moment in terms of how different businesses might use the technology. You are talking about using this technology, but is the difference about businesses in the UK that are able to use it to offer something to other businesses? Is this project that you are talking about the development of one that is using the technology to somehow create a new, emerging opportunity?

Dr Peter Waggett: The Hartree national centre was set up to provide UK industries with access to this technology. As I say, it is jointly funded by IBM and the UK Government. Customers or clients can come to us with challenges and we look at how we can make a proof of concept to deliver that challenge. The other aspect of the work that we do is very much around education in these spaces, providing courses to management as well as technical people to help them move it forward. Hartree is very much an enabling project for UK industry to come in and take advantage of the skills and resources we have.

Q14 **The Chair:** Before I move on, can I come back to you, Dr Benaich? The picture that you painted of the contrast in investment between the UK and Europe and other places is quite stark. What do you see as the

barriers or how can that be changed? What levers do you think are open to the Government to help create and capitalise more opportunities to bring that investment into the UK? We hear about how there is an awful lot of research and work going on here, but one of the things that we are not necessarily good at as a country is converting that research into commercial opportunity.

Dr Nathan Benaich: Yes, totally. This topic is quite dear to my heart. I have been working on trying to unlock a lot of the latent potential that you describe in universities. It is very encouraging to see that the Government have led an independent review of this process, which I think we will hear the news from later this year. If we set up a great engine to translate inventions from universities into companies, that will go a good way towards making this country attractive for entrepreneurial academics who want to develop technology and companies—and want to stay here.

The second thing to say is that leadership in science technology is not cheap. It costs a lot of money; it takes a lot of time. Everybody has to be bought into this. It is not only a money thing but a culture and risk thing. I think we are making strides in that direction, but we cannot get there through piecemeal amounts of money that are thrown around left, right and centre on projects.

The other aspect is attracting venture investors to invest in this country. We have gone through an unfortunate situation of a lack of political stability, which leads to lots of variations in policy and is not very good for building long-term confidence. The other aspect is that talent is incredibly important for this domain in particular. You could probably enumerate the number of people who are skilled at machine learning—and, even more specifically, language models—in thousands. A lot of those people are attracted to places where everybody else is working and they can be around their peers. That is increasingly in the US, not in Europe, so making a more concerted push to attract those individuals to this country will be very important.

I think there is a change from an overt hostility towards technology and towards the idea that it is good to be building technology companies—that it is good to be successful and not a bad thing. Flawed legislation, such as the Online Safety Bill, or various issues with government being quite slow at adopting new technologies certainly do not help, despite a huge willingness of tech entrepreneurs wanting to serve government. This is one of the emerging areas.

The Chair: Do you have a specific example of that?

Dr Nathan Benaich: Yes. There was one around a proposal from the Home Office to slow security upgrades to smartphones, which suggested an endemic hostility towards updating technology. It seems a bit bizarre. We seem to be obsessed with AGI potentially hacking the NHS, where it definitely does not need AGI to be hacked. I also recall that, a couple of years ago, DeepMind had this ambitious goal to innovate the NHS using AI. Everybody was very excited about it—certainly I was, and a lot of my

peers were—and what we got a couple of years later was nothing to do with AI. It was basically a task management app for clinicians.

The reason is not that the AI is bad. The reason is that the infrastructure in which we are trying to apply advanced AI systems is just not up to speed. It is like trying to modernise a castle; you have to rebuild the castle itself. That takes a lot of time, money and investment, to the point that science tech leadership does not come for free.

The Chair: I will have to move on because I am conscious of time, but is what you have just described unique to the UK? Is it not something that you would see in other established countries that are dealing with legacy infrastructure, as opposed to building completely from scratch?

Dr Nathan Benaich: It could be. Certain countries are in better situations than others with regards to infrastructure, but if we abstract away and say, "Where do the best technology entrepreneurs build, and where are the tech companies building products that we use every day?", those are in America. At the end of the day, you have to think from scratch and try to swing big to overcome the position you are in. I do not think we have a choice, really.

The Chair: All right. This is the sort of debate we could pursue and spend the whole hour on, but I will move on.

Q15 **Lord Kamall:** I want to touch on some of the things that Peter Waggett said and then ask Dr Webster and Dr Waggett to respond. Let us talk about how companies will integrate these tools into their operations, including customer-facing operations. I know that Dr Waggett spoke about customer service. Given that as politicians, we are always talking about what can go wrong—because people write to us when things go wrong—first, how do you anticipate large language models being integrated into the operations and the customer-facing aspect of your businesses? Secondly, who is responsible if algorithms start to inform decision-making or even take decisions themselves? Could I start with Dr Webster, then Dr Waggett? Dr Benaich or Mr Marconi, please feel to come in if you think it is relevant.

Dr Zoë Webster: Thank you for the question. We can see many ways this technology can be embedded. In terms of why we might embed, as you say, it could have a huge impact on allowing us to reimagine customer services to provide much more engaging services and interfaces to customers to allow them to do what they want to get the service they want. Whether that is the look and feel, whether that is contact centre agents or whether that is what is presented to them when they come to our website, we can get a lot more creative and give them much more of what they want. We can also reimagine the colleague experience: we are a big company, so we need to think about how to serve our own colleagues as well. No one has to start with a blank sheet of paper anymore when they are looking to design something for the customer, to write emails or to summarise a meeting. That can be done to some extent, so that gives advantages there.

Within our own operations, obviously it touches on network operations but also software development. A lot of the money we spend is on creating code, so we are piloting the use of some of this technology to speed up code development. Critically, all those things are decision support rather than decision-making, so there is a human there in the loop: for example, checking the code before it goes into use anywhere.

There are a number of reasons why we might embed it. As to how we embed it, there are several ways. Some of it will be fairly hidden in some of the applications and big systems that we make use of. Any company, for example, would have a customer relationship management system—a CRM—so this technology will crop up there. We will see more use of this technology embedded within other forms of applications and services that we might see or be using in the business. We will be using LLMs ourselves and building applications for the customer, or for colleagues or for our operations built on top of those, so that will be a lot more transparent for us. There is the use of APIs. We can access these models and again use them for decision support or to generate content that we can use in our operations or for customers.

Critically, at the moment we are piloting in many areas to understand how to embed this well to get the benefits that we think there will be, and partly to understand the cost of it, because that other part of the equation is a bit unclear. In doing this work, what is key to us is that we have responsible tech principles which we operate by, so we see AI more generally in that context. When it comes down to who is responsible when these decisions are taken, we keep that in decision support for now. If we get into some applications where the human has a very light touch, then we use our responsible tech principles to make sure we are open, accountable, transparent and fair in what we do.

Lord Kamall: It is interesting what you said about decision support. Does the buck stop with a human in decision support?

Dr Zoë Webster: Yes, a human is accountable. We have an AI standard and guidance for people—anyone in the company developing this technology or using it. It is clear that accountability should be with a person.

Lord Kamall: Before I turn to Dr Waggett, I was very interested in what you were saying, Dr Webster, about using AI in academia. I can very well see using AI to do my literature review, which is quite often half the paper, and identify the gaps. Dr Waggett, could you answer almost the same question or complement what Dr Webster has said, particularly about how you see it being integrated into operations and the customer-facing ones, but also on responsibility, particularly when things might go wrong?

Dr Peter Waggett: IBM as a company is over 100 years old. We have been introducing new technology to the business world for years. There are two aspects to the answer I will give. First is how we do this stuff internally; the second is how we take it to market through projects such as the Hartree centre. We do not release technology to the public unless we fully understand its consequences in terms of ensuring there are

guardrails—that there is proper accountability and governance on the set-up that is on there.

Currently, we have a product set that we are marketing in this space called watsonx. At every level, from data ingestion through to the model development and deployment, we have all the monitoring built in place and we make sure this is at a point where we can take it out. Those internal lessons, which we have learnt through the Hartree centre, are helping UK industry to take them on board.

We have a series of programmes under that centre. The first is Explain, which is basically training. That training is for technical people, but also for managers. Equally important as technical people are the managers, so that they understand what they will need to take on board as we bring this stuff in. Once we have people who have reached a level of understanding, we do what we call an Explore project. That is a business and technical proof of concept. In that instance, we work with the client in a very collaborative approach to make sure that we get something that maps what they are expecting, but also what their responsibilities are that come through that. Finally, we can do what we call an Excelerate project, where we help move it into their business. When it comes to Excelerate, we have a plethora of things. Some companies may come to us and say, "Look, we have done the Explore. We really understand it, we get it. We want to take this forward", and that is fine. Others are saying, "Hang on a tick, we still have concerns. How do we move it forward?", and there we have a much more collaborative set-up.

Our people—whether that is people from IBM or from the Science and Technology Facilities Council, which is the bit of the UK Government that is involved with us on this programme—can effectively hold hands when putting this stuff in place. We believe it is very comprehensive. It is built in from the beginning. We were one of the first companies that had an ethics board; I think we were the first in our industry that had a chief privacy officer. All these things are built into our processes and that is what we are hoping to roll out with the Hartree centre.

Lord Kamall: Very quickly, do Dr Benaich or Mr Marconi want to add anything?

Dr Nathan Benaich: Maybe just my two cents. It is still very early days for the deployment of this technology. Clearly, we have some large model vendors and a lot of companies built on top of that. We need to consider carefully who owns the liability of these systems, particularly as the food chain for building on top of everybody can be quite long. Model developers, who are like the bedrock service providers, need to make sure that the data they capture is ethically captured: that they have the rights to it, that it is free from biases, that it cannot be adversarially attacked et cetera. It is a bit of an open question as to how far downstream they can be held accountable, especially if one takes one of these foundation models and fine-tunes it maliciously. Is the base-model vendor liable for this bad behaviour that probably voided its T&Cs? There is a bit of grey zone in accountability—

The Chair: We will come to accountability and responsibility a bit later

on, yes.

Lord Kamall: Did you want to add anything, Mr Marconi? There will be plenty of other questions, I am sure, from my colleagues.

Francesco Marconi: Yes, I am sure I will have the chance to touch on some of them.

Q16 **Baroness Harding of Winscombe:** I just want to follow up with Dr Webster. You talked about running pilots. I am interested in how you are evaluating the performance of the large language models that you are piloting, particularly on whether they are hallucinating and how, going forward, you will provide the quality control and quality assurance as you move from pilot to rollout.

Dr Zoë Webster: Thank you for the question. Hence our reason for focusing on decision support: because we are more than aware that there are issues with hallucination. In the case of co-generation, for example, one of the rollouts that we are doing from the pilot is to capture for the number of users involved—and that is a carefully curated set of people involved in the pilot—how many lines of code they are accepting and committing from what the LLM-enabled capability was providing. From what they were given, we can capture how many of those they checked and deemed acceptable for use. From that, we can also work out which ones were not, because they were either calling on a library that does not exist or were just not appropriate. We are seeing high percentages of acceptance, so it is looking positive, and it bears out the evidence that seemed to be out there: that you can get an increase in productivity of between 20% and 30% from this, at minimum. We are looking at metrics that give us a very useful indication of how the person in the mix is taking on board what is being suggested.

Baroness Harding of Winscombe: Dr Waggett, do you have anything else to add on how you are evaluating the effectiveness?

Dr Peter Waggett: Yes. It is very similar to what Zoë has been saying. In particular, with the watsonx product that we are developing, a call went out to the whole of the IBM organisation to get involved in a jam to take this thing, deploy it and try to find ways to break it. We are pretty good at breaking some of these things. We capture all those lessons that we learned from trying this stuff internally, and put them back into the system to make sure that when it goes out to customers, it is fit for purpose.

Lord Young of Norwood Green: You said you have an ethics board. What are the key ethics that you find?

Dr Peter Waggett: When we are putting these systems together, we are looking for accuracy, trustworthiness and whether we are getting repeatable results coming through. It has to be explainable and adaptable; it has to be transparent. The transparency is the key thing that will help with this technology to make sure that we are not using black-box approaches to some of these capabilities. As we put together

our testing schedules, we test against the criteria to make sure that we can do it.

Again, it is a slightly hackneyed example, but as we are over 100 years old, 100 years ago business machines were scales. The analogue I would give is that if we had a set of scales and there was a screen around the pan that has the weights on it, then if I came in to buy something I would not trust what has gone on with that. We are trying to understand how we can make this available and applicable to people as they start to use the technologies.

Lord Young of Norwood Green: Just briefly, Dr Webster, I have to declare an interest here. I worked for BT over many years and negotiated with them. You have declared—or Philip Jansen has—that about 10,000 jobs will be replaced. I wonder, because this is a big transition, whether you are negotiating with the unions and how you feel that your morale is being sustained in such a transition.

Dr Zoë Webster: That was as part of a transition right to 2030, so there is time. I think we are engaging with our workforce. As I say, we are trying to transform their experience as well and show how technology can be useful to them to get their work done and make that better. A lot of this is about colleague engagement, hence having them involved in some of these pilot rollouts. It gives them the opportunity to try these things out and find out how they work for them, and to give us feedback as a wider team on what works and what does not.

In reality, what we are finding—and I think what others are finding—is that this technology is not magic. You do not press a button and out comes a beautifully crafted response. There is still the need for a lot of expert human interaction. There needs to be someone breaking down the problem in a way that makes sense by using this technology and then putting something back together again that meets the need. It is not that everyone will go suddenly. We do see this augmentation role. This technology will augment people in their roles and make it slightly easier for them to do some tasks, but there will also be additional roles. The prompt engineer is the classic example. We did not know about prompt engineering a year ago and now we have people with—

The Chair: What engineering, sorry?

Dr Zoë Webster: Prompt engineering. It is people who understand how best to present prompts or queries to something powered by an LLM to get the best response out. That is one example of a new role that we have started to see and there will be more, as we see in most technology advances.

Going back to your question, it is more about that engagement and getting people involved in those pilots—getting their feedback—but we know that we will still need expert interaction. This is not a magic technology that can replace everyone.

Q17 **The Chair:** Thank you. Just before I move on to Lord Lipsey—where, Mr Marconi, we will be asking questions of your good self—I have a couple

of follow-ups on this section. Picking up from what you were just saying to Lord Young, Dr Webster, how you would answer this question, which I think troubles a lot of people? If this is black-box technology where nobody understands what is happening, how can people be confident in adopting it? Is that what you have just said and how you would answer it, or would you say something else?

Dr Zoë Webster: This is a really important question. To a certain extent, they are black box. People understand the approach and the principles behind it, but I think even the engineers of some of these technologies are surprised by how well it can work, and that hinders adoption to a certain extent.

Going back to a comment earlier around liability, the emphasis to date has been on the big developers of some of these foundation models and, understandably, on the end user. However, the deployers in the middle often get forgotten. There are concerns that we will be held accountable, or could be, for issues with a foundation model where we have no idea what data it was trained on, how it was tested and what the limitations are on how and when it can be used. There are open questions and that is a limiting factor on adoption. We know that we need to adopt AI more generally in the UK.

The Chair: I do not know whether you or Dr Waggett are best placed on this—I will come on in a moment to Dr Benaich—but in terms of encouraging other businesses to adopt this technology and its commercial opportunities for them, is this one of the biggest barriers?

Dr Zoë Webster: It is one of the biggest barriers. Uncertainties around cost is another, but this is a very nascent technology area. The big platforms are still trying to work out exactly how they are responding. There are still new models and forms of models. The ways of engaging with models are still being developed. For a company to work out how much to budget against that is actually very challenging, and that is another barrier alongside—

The Chair: We will come back to that later on when we get to a bit more on responsibility. I have one final thing before I move off this question. Dr Waggett, coming back to what Dr Benaich said earlier about the problem of legacy infrastructure, how do you at IBM address that? Are you building or developing something that means you are precisely conscious of legacy-type infrastructure and it being a barrier to adoption of this? It just occurs to me that if this is a main obstacle to taking full advantage of the technology, is IBM particularly concerned about and focused on it?

Dr Peter Waggett: One of the key things that IBM is pushing is what we call the hybrid cloud approach, which takes account of the fact that we are having to integrate legacy elements. However, before anything goes into the systems we need to understand what is going on.

I have been working in this industry for a while, and one thing that I learned early on was that it is vital to understand the data that you are taking into a system. Do not just take anything at face value. In my

personal case, my background is in rocket science. I used to use the ozone layer as a calibration constant. My sensors would look at the ozone layer; I looked at it through that system for probably about three years while I was doing my PhD. As I only found out that there was an ozone hole when somebody announced it, I just sat there thinking, "Why did I not spot it?" What had I missed? As it turned out, the assumption had been made in the database that if the data was not constant it must be wrong, so throw it out.

In that instance, I learned very early on that you must understand what is going on in the data. Some practitioners will say—particularly data science practitioners—that about 80% of the job is understanding the data that you are going to ingest before you come to do anything with it. So yes, we are happy to take on board legacy data, but not without it going through audit and by understanding what we are putting into the systems.

Q18 **Baroness Harding of Winscombe:** In what you have just described, does that mean that there is a fundamental flaw in the way that the LLMs that have been let loose on the public are constructed because, by definition, none of us have the slightest idea of what data was used to construct them?

Dr Peter Waggett: No, that is not what I said. If I gave that impression I was not—

Baroness Harding of Winscombe: Then I misunderstood.

Dr Peter Waggett: Yes, sorry, what I was saying in that instance was about the dataset that I was using. You need to understand it before you can understand how to apply it, so—

Baroness Harding of Winscombe: Is one of the biggest problems with large language models that none of us know what the datasets are that have gone into them and, therefore, if I follow your logic, we cannot use them sensibly because we do not understand what has gone in?

Dr Peter Waggett: No, again, my apologies if that is the impression I have given you. It was not what I meant to convey. What I am saying is that we need to understand those datasets that have gone into systems that we put together, and we do. In the systems that we put together internally, and those that we deliver to end clients, we do understand them and the data that is put in place.

Baroness Harding of Winscombe: I do not want to keep going. I know the Chair is keen to move on.

Q19 **Lord Lipsey:** Mr Marconi, like me, once upon a time you were a journalist. Will there be any journalists left in five or 10 years' time?

Francesco Marconi: I hope so. In the context of this conversation on large language models, the potential application to journalism in the news industry is that it can enable two fundamental things. One is helping newsrooms to expand news coverage. The second is overcoming the human capacity limits, or restraints.

To try to answer your question a little more directly, the way to think about the impact of LLMs in journalism is to think through the different values of the news process in newsgathering, production and distribution. While LLMs are good at brainstorming ideas or finding interesting story angles, they are not fully reliable in producing factual data. That is one of the fundamental roles of a journalist: to research those data points and so on.

When it comes to production, starting with the assumption that we do have those interviews—those data points—transforming that raw data into a story is something where a large language model excels. If we think about functions that are related to repurposing information, curation, and summarisation, those roles are perhaps at the highest risk. Then the third area of impact will likely be the distribution of this information.

Today, we are used to consuming news on our mobile devices and perhaps we find news stories on social media or search engines. It is likely that the emergence of language models will lead to a change in the behaviour, where the consumer expects to converse with information, rather than passively consuming it.

At the end of the day, for this hopeful and positive vision of LLMs to come to fruition as it relates to the news industry, there are two fundamental challenges to be addressed. One is the recency constraint. There is a data cut-off, in that the models are trained until a certain point in time. If we ask for information on an event that occurred last month, if that language model is not connected to a real-time database it is useless for journalistic research. Therefore, there is a recency factor and then the accuracy, which is what we have been discussing, where there is a lot of room for improvement. A lot of techniques are being developed to increase the level of accuracy, which is fundamental when we think about the news industry and the application of AI.

In the long term, as we solve the issues of accuracy and those of a high-cost structure, it is very likely that language models will be an integral part of any newsroom. I like to think of it as almost like a free intellectual energy that allows journalists to focus on high-impact areas in automating certain more mundane, repetitive tasks.

If I may mention one more thing, beyond the discussion around the technical limitations, the accuracy and so on, there are also business and legal challenges that the news industry is reflecting on. Specifically, there is the fact that AI models rely heavily on news data to train their models, in addition to books and other sources, but publishers are not necessarily compensated as the providers of this data. The reason I bring this up is that for us to have a long-term, sustainable AI sector we need an equally sustainable news sector because it is the source of high-quality information.

There are these two aspects to consider. One is the accuracy and technology considerations, but also how we think about the business and legal frameworks that exist today.

Lord Lipsey: I understand the point about accuracy, although I must say that human journalists are not very good at it either, which is why some of us started the fact-checking organisation Full Fact. However, what strikes me as a much greater weakness of the machines is the prose they produce. I do not know if you saw that long piece in the *Guardian* the other day when it asked an LLM about Proust. It came out with something that was between dull and gibberish. We have had experience of that in this committee with some of the things we read. They do not seem very good at prose and that seems to be quite a big inhibition.

Francesco Marconi: As I mentioned, the root cause of these so-called hallucinations that you were referring to is often related to the two limitations that I mentioned of recency, in a lack of recent data for these models, and accuracy. There are ways to address some of those issues and there is a nascent field in the news industry, which is computational journalism. It is merging techniques of artificial intelligence and machine learning with editorial-based practices. The question is: how do we address these inaccuracies? How do we fact-check the machines if they are producing information at scale?

If you are a news organisation and you are applying a language model to your news-gathering process, there are a few things being done. One is the technique which is called fine-tuning, where we take a pre-trained large language model and expose that model to news data; for example, articles about a particular subject, such as climate or the environment. What that allows us to do is to emulate the writing style and have a better grasp of the language that appears in those coverage areas, but it does not necessarily guarantee better accuracy on those topics. The reason for that is that LLMs recognise the statistical relationship, so what the likelihood is of a certain string of words appearing in a sentence. They do not necessarily understand the words' meaning. Perhaps that is where it stemmed from in the *Guardian* story that you alluded to.

However, there are promising solutions and one of them is the field called retrieval-augmented generation, or RAG for short. In simple words, it is a new approach to connect a large language model to a reliable database of fact-checked information. Rather than relying on this language model to do the research and the articulation in the writing of a story, we instead rely on these external databases for facts. The LLM's role is simply to translate that data into coherent language.

You could imagine this methodology being applied to real-time news coverage when, for example, we are talking about government agencies as they relate to health services and so on. They were collecting public health data on the spread of Covid and so on, so connecting those databases that are accurate and factual with LLM means there is a lot of potential.

Lastly, I would like to highlight that often the news sector and its journalism is regarded as a spectator in the context of AI, but I believe that journalists have an active role to play in the development of AI. Not only is there a wealth of data that everyone wants to train these models

in; they also possess the ethical frameworks and the fact-checking workflows to implement them into the model. What that means is that rather than simply training the models on news data to teach a machine how to write, we can also teach the machine editorial principles. This is the concept of editorial AI or constitutional AI, where we have the machine abiding by a set of principles that are defined by journalists.

The Chair: Thank you. We have a brief supplementary from Lord Young, then we must move on because we have very little time.

Lord Young of Norwood Green: Briefly, your interactive journalists are interacting with social media all the time. How do you incorporate that into your large language models?

Francesco Marconi: Could you please clarify on integrating? Do you mean how to leverage language models to verify social media?

Lord Young of Norwood Green: Yes.

Francesco Marconi: The broader theme is that of augmentation. The work that a journalist can do manually can now be scaled through machines, so the idea of using social media almost as a real-time pulse of society in terms of what topics are being mentioned. Are there particular shifts in sentiment? I believe the application of language models to social media can be very promising as it relates to creating a new type of newswire to alert journalists and get them more in touch with the general public, and doing so in a more efficient way.

Q20 **The Lord Bishop of Leeds:** I would like to go back to a question you have already touched on in relation to governance. It is on liability. As some of these tools become more widespread, where do you think the balance between responsibility and liability should rest between the upstream developer and the downstream deployer? I would be interested to hear each of your perspectives from your different roles.

Dr Zoë Webster: Thank you for the question. I do not think it is where it needs to be right now. As I mentioned earlier, as a deployer it feels like a lot of the liability rests with us. If something was to go wrong and if that was down to the model—and we know there is or can be bias in many of these models, based on the data they are built with—it feels like that would land on the deployer's lap only.

I think we have an asymmetry of information, though. We do not have complete information about the models that allow us to make good decisions about which models we can use and how we can use them. At the moment, we are largely trying these things out safely to see how they work. There needs to be a balance between that return on the investment that people have made and will make into deploying these models and making sure that the liability is fairly assigned. At the moment, it feels like it is squarely with the deployer and we do not know everything we need to know about these models to be assured to adopt these things safely, so there needs to be a shift.

The Lord Bishop of Leeds: To a layman like me, I cannot see how that would ever change because you have also described how some of the

aims are becoming ever more complex or longer. How will that ever change?

Dr Zoë Webster: There may be a role for some form of regulation here, even just by regulating the use of something like model cards, which IBM and others are using, in some format. It is regulating so that the supplier of a model provides much more information to people who might deploy it about how that model has been derived, how it has been tested and what possible constraints there should be on its use.

The analogy would be that if I was to start up a taxi company, I have some form of warranty or assurance from the company that has built the car. I do not need to know exactly how it works but I have assurances that it is safe to drive, as long as I drive it within certain bounds. At the moment, we do not have that within the context of LLMs, for example.

The Lord Bishop of Leeds: Like the VW and diesel test.

Dr Peter Waggett: I agree with Zoë. Both developers and deployers must be made accountable; both bear responsibility and neither should be immune from liability. That is particularly why, when we say that we work closely with companies, that collaboration is needed to make sure that we have the right solution coming out. Because neither side has a full story in terms of a particular user case or a particular application, it needs to be collaborative.

Dr Nathan Benaich: I am definitely not a legal expert on this, but my two cents would be that it is unlikely to sit in one part of the value chain. It is important to get it right because from an early start-up company's perspective, if you develop a product that uses AI and your customer says, "I want unlimited liability insurance, in case your thing does something I am not happy with", that could just blow up your business. We have to put caps in or figure it out somewhere.

I cannot think of any other industry in the world that would place all the liability on the developer of a technology for all its downstream uses, so maybe that is not the right place to look at either. If we do not accept a degree of risk or personal liability, we just stifle technology progress in general.

Francesco Marconi: Yes, I agree with what was just said, but there is an outlier here. In some cases, as it relates to news generation, there is a blurring of boundaries between the upstream developer and the deployer of AI. If we think about the context of generative AI search, where a technology company is the developer of the model but then also implements those models into search results, rather than showing a list of links of news sources it is picking and choosing different facts or passages from those news articles and presenting a summary. So when companies in that scenario both create the AI and deploy it, there is heightened responsibility and accountability.

In journalism, the human editors are accountable for mistakes, but how do we do that at scale if there are possibly millions, if not billions, of news summaries being generated every day? How do we vet and determine accountability and liability at scale? In this realm, that is

where the fields of ethics, algorithmic transparency and journalism itself come into play. That means embedding these models with checkpoints for validation in terms of trustworthiness, transparency, accuracy and so on.

The Lord Bishop of Leeds: Presumably, even they can be falsified or worked in such a way as to guarantee reliability when they are not reliable.

Francesco Marconi: I think that is uncharted territory and—

The Lord Bishop of Leeds: That is partly why I am pressing the point. It seems to me that there needs to be a fundamental humility here: when you are dealing with the length of these value chains, and their complexity, there is no way you can guarantee.

Can I just press a further question? We have been interested in the difference between open and closed-source models. Do you think the burden of liability shifts depending on whether the model was open or closed?

Dr Peter Waggett: You need to assess it and work it through. I do not think it is possible to say one thing or another. Unfortunately, it is the consultant's answer: it depends.

Dr Zoë Webster: I would say it depends on use. If you go back to its context rather than the technology, it is about how it is put to use. As Peter said, it depends on the circumstance. Both have pros and cons. It is early days and we will see how that pans out. Obviously, openness gives a level of transparency that helps with that information asymmetry, but then it potentially opens up to security issues. With closed, again, you have the limitations and the costs but, potentially, more support available. We will have to see but, in terms of liability, it comes down to the context behind its use.

The Lord Bishop of Leeds: Do you have anything to add?

Francesco Marconi: I think that open models can alleviate some of the data privacy concerns that news organisations have. The way that these models work is literally that we can download them and then host or run them in house, so there is more control and a higher level of scrutiny. At the end of the day, there will never be full transparency in how these models behave. Although there is a significant difference between open source and proprietary or private models, it is still a complex issue in terms of the accountability and liability of the models.

The Chair: Thank you. Is this a brief comment, Dr Waggett?

Dr Peter Waggett: It is just a comment, really, because I think we are in violent agreement but we do not seem to have reached it. Legacy datasets can be used to train large language models. IBM will not use a large language model where we do not understand the data that has been used to train it. I think that helps to understand what we are trying to say.

The Chair: I think Baroness Harding was more talking about LLMs

generally rather than IBM.

Baroness Harding of Winscombe: Yes, I am more concerned about—

Dr Peter Waggett: What I was describing was how IBM does it.

The Chair: Yes, it was a difference in what we were talking about.

I thank all four of you hugely for your testimony today and for your time, particularly those who may have travelled distances to be with us as well. I am very grateful.