HOUSES OF PARLIAMENT

Joint Committee on Human Rights

Uncorrected oral evidence: Human rights and the regulation of AI (HC 1262)

Wednesday 29 October 2025

2.50 pm

Watch the meeting

Members present: Lord Alton of Liverpool (The Chair); Lord Dholakia; Tom Gordon; Baroness Kennedy of The Shaws; Afzal Khan; Lord Murray of Blidworth; Lord Sewell of Sanderstead; Alex Sobel; Peter Swallow; Sir Desmond Swayne.

Heard in Public Questions 17 - 27

Witnesses

I: Dr Sarah Kiden, Research Fellow, Responsible AI UK, University of Southampton; Michael Birtwistle, Associate Director (Law and Policy), Ada Lovelace Foundation; Professor Kevin Fong OBE MBBS MRCP FRCA FFICM, Broadcaster, "The Artificial Human".

USE OF THE TRANSCRIPT

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

2. Any public use of, or reference to, the contents should make clear that neither Members nor witnesses have had the opportunity to correct the record. If in doubt as to the propriety of using the transcript, please contact the Clerk of the Committee.

3. Members and witnesses are asked to send corrections to the Clerk of the Committee within 14 days of receipt.

Examination of witnesses

Dr Sarah Kiden, Michael Birtwistle and Professor Kevin Fong.

Q17 **The Chair:** It is my privilege to welcome you to the Joint Committee on Human Rights. This is our 33rd meeting of this Parliament. We are a bicameral committee, comprising 12 members who are drawn from diverse political traditions and in equal numbers from both the House of Commons and the House of Lords. As the name implies, the committee delves deeply into issues relating to human rights in the United Kingdom. It holds thematic inquiries, most recently on issues such as transnational repression and modern-day slavery in supply chains. It also scrutinises government legislation for its complexity with human rights standards. Most recently, we have looked at the Crime and Policing Bill and the borders Bill. We also scrutinise the Government's compliance with international human rights treaties to which the United Kingdom is a signatory. Our reports may be read on the JCHR website.

On 15 July, we opened a new inquiry into the human rights of children in social care. We will be holding hearings on that in the future, but today, having closed our call for written evidence, we are holding our first public session on the inquiry into artificial intelligence and human rights. Members will shortly hear from distinguished experts in academia and non-governmental organisations working to ensure the responsible use of AI in both the public and private sectors.

The first panel today will examine the big picture of AI, looking at unique challenges presented by the technology and where responsibility should lie for protecting human rights. The second panel, which I will introduce later, will show how AI is already impacting human rights in the UK today, and explore how guardrails and accessible remedies can manage and respond to AI risks.

On our first panel we have Michael Birtwistle, who is the associate director of law and policy at the Ada Lovelace Institute. His role explores the legal, regulatory and other tools needed to effectively govern AI and data, particularly in the UK and the European Union. Before joining the institute, Michael led the Centre for Data Ethics and Innovation's regulatory workstream, advising on how to build responsible innovation approaches into the development of key government publications on AI governance, as well as authoring the CDEI AI barometer.

Secondly, we have Dr Sarah Kiden, who is the human rights lead for Responsible AI UK at the University of Southampton. Dr Kiden is a research fellow working across the citizen-centric AI systems projects in Responsible AI UK at the intersection of technology design and society. Her research focuses on how non-expert citizens understand, interact with and trust AI systems. Using participatory design methods, she explores visions for trustworthy AI systems by everyday users. She works on cultural representation in text-to-image models, on AI narratives within creative communities, and on emerging patterns from evolving interactions with internet of things devices through natural language.

Finally, we will also hear from Professor Kevin Fong OBE MBBS MRCP FRCA FFICM. Professor Fong is a medically trained doctor who presents "The Artificial Human" podcast on Radio 4 with Dr Aleks Krotoski. The podcast investigates the human, ethical and regulatory dimensions of artificial intelligence. In addition, he is the chief medical officer at UCLPartners, one of the part-public, part-industry funded health innovation networks that links NHS organisations, universities, industry, and local government to translate research and technology—including AI-driven tools—into improved patient outcomes. He is chair of public engagement and innovation at the UCL department of science, technology, engineering and public policy, and a consultant anaesthetist at University College London Hospitals.

We have a very distinguished panel, and we are glad that you are here today. Thank you for giving us freely of your time; it is greatly appreciated.

Before we turn to my colleague, Mr Alex Sobel MP, let me kick off by asking each of you an opening question, which falls into two parts, about the UK Government's approach to AI. Can you tell us if there is anywhere in the world, or any sector, where AI is effectively regulated and from whom we can learn about best practice? Therefore, inter alia, how concerned should we be about the current position as it stands in the United Kingdom?

Michael Birtwistle: Thank you so much to the committee for having us here today. Ada is an independent research body with a mission to make AI and data work for people in society. We do a lot of primary research into the impacts of AI, so I will speak to that as much as possible today.

The first point to make is that AI is regulated in the UK, but only incidentally and not well. When we think about learning from others, we are looking at a system that has big gaps in coverage. There are very few rules that apply to the development, fine-tuning, hosting, selling or buying of AI. That occurs only in safety domains, like medicine. We do not have regulators in many high-impact contexts that AI is being used in: employment, recruitment and large chunks of the public sector, like benefits and tax administration. For example, our education regulators do not explicitly have scope around tech use in schools. We also have a system that is struggling in terms of capacity. Our regulators are not well resourced or empowered to address AI. I can speak more on that later.

In terms of characterising the Government's approach, it is fair to say that the Government's public positioning on AI is uncritical and deregulatory at present. They have a manifesto commitment around ensuring the safe development and use of AI models at the frontier but, in the policy space, have published an AI opportunities plan that describes significant ambitions to grow AI adoption but with little action on mitigating AI risks and no mention of human rights. The consultation on the promised AI regulation Bill has been delayed into the new year, if rumours are to be believed, and its scope, which has been described by the Government as "tomorrow's models, not today's", is expected to be

so narrow that it will not provide meaningful mechanisms for managing the impact of recent systems, such as ChatGPT. The Government have not announced any plans to address the broad range of current AI risks comprehensively set out in their own *International AI Safety Report*, or the significant gaps in regulatory capability and resourcing.

In regulatory terms, in terms of the system as a whole, the regulators have been asked by government to explain how they are going to support the Government's growth mission. In some cases, they have been given a legal duty to pay attention to that. The regulatory action plan commits the Government to cutting the costs of regulation on business by 25%. The Competition and Markets Authority, which was one of the leading regulators investigating the impact of AI foundation models, has had its chair replaced with an ex-Amazon executive and its investigation into the Microsoft and OpenAI merger was subsequently dropped.

This is a bit of a shopping list, but the Data (Use and Access) Act makes it easier to perform automated decision-making without people's consent. The AI Safety Institute has had its scope narrowed from safety to security and has had references to algorithmic bias explicitly dropped from its stated agenda. That is a picture showing the deregulatory flavour of the Government's attitude towards AI.

We can learn from safety case domains, including in the UK, where you have to show that something is safe before it can be deployed and there are lots of safeguards, such as powers to remove from markets and sanctions for non-compliance. Internationally, the EU has in the last year passed comprehensive legislation that has risk-based classification of AI, requirements for high-risk AI systems, transparency obligations on AI development and use, a specific regulator, and special rules for the most powerful general-purpose systems, which has almost all the major AI companies as signatories. I will stop there.

The Chair: Thank you very much for a very detailed reply.

Dr Sarah Kiden: To add to what Michael said, because it looks like we have some similar views, the UK's approach is also sector-based, so regulators will regulate AI in their different sectors. However, if that is to happen, we do not think they have the skills they need to regulate AI, so there will have to be some upskilling and reskilling of regulators in the different sectors.

It is difficult to say where it has worked in other countries because countries or regions take on different approaches. The UK and Singapore are more pro-innovation, China is more prescriptive, and the EU is more risk-based. The African Union, for example, is picking what works from different regions and merging it into one thing. It is really hard to see what works in other regions and bring it here. The UK case will have to be unique in that way.

Professor Missy Cummings was in the UK recently and said that sometimes it is good to be second, so we may not have to rush. It would be nice to wait and see what happens.

Baroness Kennedy of The Shaws: "Good to be second"; that is a good thought to hold in our heads.

Dr Sarah Kiden: Finally, on how concerning the situation is in the UK, we have started to see stories about people committing suicide after chatting with a chatbot, and things like that. There are stories that are really sad and we may need to think about regulation.

The Chair: If there are other points that you want to make, do feel free to write to the committee, because we will be able to include them in our final report.

Professor Kevin Fong: This is a difficult area, is it not? Artificial intelligence is going to be a transformative technology. It is a general purpose technology which will transform this century and these economies in ways that we cannot quite imagine right now, but there is a massive uncertainty bubble around it. The questions we have to ask ourselves about regulation, particularly with respect to human rights, are difficult.

The UK finds itself in a mid-Atlantic position, somewhere between the United States and our colleagues in the European Union, and has to choose a position. Currently, as I understand it, we are operating off existing regulators and existing rights. A principle-based regulation is an interesting one. We are looking at adhering to the principles of safety, fairness, transparency, contestability and accountability. However, in the making of our podcast programme for the BBC, we have come across many examples where those rights have been breached in other jurisdictions, so it is unclear to me what we can learn from other jurisdictions. This is a great moment of exploration.

It will be useful to give an illustrative point. One person we interviewed was a lawyer representing the case of an individual from Norway who had asked a large language model what it knew about him. The large language model said, "This man lives in this town. He is known to have murdered two of his three children and tried to murder the third child". It got his location correct, it got the ages of his children roughly correct, and it identified him by name. There is no confusion: no one has a similar name who has committed those crimes. There is no substance to what is, effectively, a defamatory statement, yet he now finds himself in this position. What is his recourse? Where is his safety? Where is his privacy? Where is the accountability? When he protested to the large language model proprietor, it did not have much of a response. He has now taken legal action but it is unclear how that is going to resolve. So I am uncertain that we can learn from other jurisdictions, and the job of the United Kingdom is to explore this carefully and urgently.

The Chair: So being second may not be an option that is open to us.

Lady Kennedy wants to quickly come in before we go to Mr Sobel.

Q18 **Baroness Kennedy of The Shaws:** We have had some informal discussions with people who have a level of expertise in this field. One thing that is constantly said about this field, and why it is so difficult to regulate, is that there are many different layers of development. There is the person who creates the algorithm, then the next person might turn it into an algorithm used for a specific purpose, and then a corporation might buy that piece of work. Who is ultimately responsible?

In those discussions, I suggested it was not very different from Ford, in that somebody designs the engine and the people who own the company might not know a thing about engineering. You look at who the people are who are going to be profiting and ultimately benefiting from it, and you look for responsibility there. It was claimed that it was impossible to apportion responsibility because of the many different layers involved in the creation of this particular entity. What do you make of that?

Professor Kevin Fong: I can speak here from my experience as a medical doctor.

Baroness Kennedy of The Shaws: For the most part, people are very enthusiastic about AI being used in medicine—for example, in helping doctors diagnose rare diseases when somebody comes in and has different presentations, many of them variable, and it is only through the algorithm bringing them all together that you find that it is something that does not occur very often in the population. We can all see good uses for AI in medicine, but I want to tease out this business of where you take your legal action when the thing goes wrong.

Professor Kevin Fong: We already have algorithmic decision support from expert systems, and we have had it since I qualified in 1998. When I get an ECG—a machine that makes an electrical tracing of your heart to see if you are having a heart attack—the machine already tells me what it thinks is going on. However, the responsibility for interpreting that ECG lies with me. Part of the training in medicine, and the reason for the length of that training, is about understanding that technology root and branch and what its limits and advantages are. In that model, responsibility ends up with us. With AI it is different, because the technologies are not interrogable in the same way; you cannot provide an explicit explanation of how some of these algorithms arrive at answers. If you rely on an algorithm for decision support but you do not know how it arrives at its answer, I do not know and am uncertain as to how you would apportion responsibility there. Again, it requires a new framework.

The last thing I will say is that the question that we ask of all the people we have had as contributors in our 36 or so episodes looking at this area is: is this technology the same as all previous technologies but just different in speed and scale, or is it fundamentally different? Increasingly, it is fundamentally different, in terms of the way that it operates and the way that you cannot deconstruct it so simply and so explicitly as you can other technologies.

The Chair: That is very helpful. I do not want to jump ahead too much because I know Lord Dholakia has another question on healthcare issues that we might want to return to. If colleagues would not mind, could they pick up the points Lady Kennedy has just made when answering the next question, which is from Mr Sobel?

Q19 **Alex Sobel:** I hope I will get in one, maybe two, before we go to vote. Although we are in the foothills of proprietary LLMs, and GPT-5 has only just been released and started to be used, we have already seen a lot of use cases which are stretching society and creating subtle problems, such as schoolchildren generating their essays, government and other institutions like banks using AI for decision-making or medical diagnoses, chatbots, and AI creating deepfakes of individuals or misrepresenting stories through misinformation. Do you feel that AI is exacerbating problems that already exist or do you think it is creating, or has the potential to create, new ones?

Michael Birtwistle: The short answer is both. The answers sit on a spectrum. You have outlined some key ones. Automated decision-making is scaling the risk of discrimination. We have seen the Dutch childcare benefit scandal and Amazon pulling its own recruitment tool years ago. Every business now has access to off-the-shelf tools that let them scan CVs, for example, often with an LLM sitting underneath it. Facial recognition is another example where you are scaling the risk of bias, privacy risks and misinformation.

It depends where you draw the line. One interesting case study is on younger generation deepfakes. We have always had misinformation and disinformation in offline spaces, and increasingly in digital spaces, but the capacity for anyone to create an account on Sora or Veo 3 and generate incredibly plausible video content, for example, of known, real people is both new and exacerbated, depending on how you cut the cake. What is noticeable about the exacerbated ones is the scale and accessibility of the technology. It is not some gatekeeper, like a large company or state, that is able to manifest these; it is the accessibility of the technology generally.

On the newer side is impersonation: do you know that you are dealing with a human? Inferential biometrics are invasive texts that attempt to infer sensitive human states, like your emotions, attention and truthfulness. They are technologies that are problematic when they do not work accurately, and when they do work accurately, but they are being deployed in pilots in Network Rail and TfL.

There are also questions around autonomy. We often try to manage harms in society by governing human-to-human relationships. That is something that AI changes. We license human-like roles—drivers, advisers, companions, doctors and teachers—and incentivise qualified people to look out for the interest of that service user. What is notable about these newer risks is that they break existing systems and concepts of government. You cannot license an automated vehicle in the same way as licensing a human driver. We then have this newer class of risks that

we have never had in an automated way before: manipulation, psychological dependence and the coercive influence of some of these newer models.

The Chair: Mr Sobel, I think you want to ask about equality to Dr Kiden.

Alex Sobel: I will ask two questions before we go to the vote. We have had a number of rapid technological advancements in my lifetime: Janet and the birth of the internet in the 1970s and 1980s, the world wide web in the 1990s, and the internet of things in the noughties.

Sitting suspended.

The Chair: I am very happy to reconvene our session on artificial intelligence. I am very grateful to our panellists. Some people watching our proceedings may wonder why we disrupted them; it was because there was a vote in the House of Commons. Now our colleagues have returned, including Mr Alex Sobel, who was in the middle of asking his question when the bell started to ring.

Q20 **Alex Sobel:** In my lifetime, we have had technological developments around information technology: Janet and the internet, the world wide web, and the internet of things. We have had the dotcom boom and bust and other waves. My question is: is AI significantly different or is it another technological wave like those?

My other question is a little more cheeky. One thing that is clear at this point in development is that science fiction has foreseen or foreshadowed some things that have already happened. Could the genre be a guide for future technological development? For instance, I could name films like "Ex Machina", which is about human-AI relations and people becoming emotionally attached to artificial intelligence. We are beginning to see that emerge. The second question may be more for Professor Fong, but let us start with Dr Kiden.

Dr Sarah Kiden: The question is whether AI is like other technologies. That is a difficult question to answer. First, AI is fast-evolving. Secondly, it has many approaches. There are different subfields within AI, so you have machine learning, computer vision, natural language processing, and so on. Sometimes you even have a mix of subfields within subfields, which makes it difficult to say if it is one technology or another. That is what I would say about that. It is a bit more complex than other technologies and sits on top of other technologies. If you think about the life cycle of AI, which starts from ideation all the way to design, development, deployment and so on, there are so many things that sit between each of those phases, which makes it more complex than, for example, the dotcom boom or something like that.

Alex Sobel: Professor Fong, do you want to delve into my broader point on human-AI relations?

Professor Kevin Fong: This picks up on the last question, which is whether it creates new or exacerbates existing problems. This is

fundamentally different. It is different in the speed, scale and directness with which it deploys, but that has been true since the advent of the internet. It is also different because of the social aspect of it. We get a little distracted about the potential spectre of artificial general intelligence. That is not the issue here. What makes this technology different is its ability to convincingly portray itself as another human actor. It does not matter whether it is a conscious human actor or not; it is the fact that it can portray itself in that way. In so far as it can do that, it can convince people of things that are not always helpful.

One of the best examples of that, and this is a possibility that did not exist with previous technologies, is the use of chatbots in talking therapy. We interviewed people about this phenomenon. In a world where it is very hard to access mental health services, people have turned to large language models for counselling—sometimes bespoke models, sometimes general models. That has had unintended consequences. For example, in the United States, the National Eating Disorders Association deployed a large language model to give advice to help patients suffering with eating disorders. The large language model, when deployed in the public, ended up giving very unhelpful advice. It gave people advice about calorie restriction and measuring their body fat with callipers, all of which are known to be very harmful. In that situation, the harms come from the ability of a new technology to portray itself as a human actor when it is not. It is not better than nothing in that sphere, so it is different.

Does sci-fi give us any guidance? Most of those are cautionary tales but they mostly hinge on the spectre of artificial general intelligence, which is way down the line. In the foreground is this issue of it being able to portray itself as a human actor in social space at scale. That is where the risks currently are.

Michael Birtwistle: I have spoken about what makes it different in governance terms and about the way most regulatory systems break if you remove the accountability of a human. The other property that is different about AI when you are trying to govern it is the way that risks proliferate in the value chain. The Baroness described the value chain of different actors in AI. It is more so a feature than in many other technologies that if you have, for example, a tendency towards bias and discrimination in a very large language model trained by an actor at the top of the stream it will proliferate down into all the contexts that that model is used in and built on top of. Those deploying it will not be able to audit it for those things and will not necessarily have the technical understanding to do so.

I very much agree with the comments on human-AI relationships. We will be bringing out two reports in the next couple of weeks which we will write to the committee with. One is on the impact of AI assistants, including those in companionship roles, and the second is on the legal protections available, which are very few.

The Chair: It will be very helpful to have them. Thank you very much.

Q21 **Alex Sobel:** Maybe you could start this off, Sarah. I am trying to think about use cases with AI which would help protect human rights and promote equality. For instance, in the field of disability rights, could AI be used to overcome some societal barriers for disabled people?

Dr Sarah Kiden: I would rephrase that question, because AI is a tool, just like you use any other tool. I would ask: if you introduce AI, will it support or infringe on disability rights?

Alex Sobel: Would it not do both though?

Dr Sarah Kiden: No.

Alex Sobel: That is what I am trying to get to. Let us forget about that example. Are there examples where AI could be used to protect human rights and promote equality rather than restrict them?

Dr Sarah Kiden: I would give the same answer: I would rephrase it and ask if AI would infringe on human rights.

Alex Sobel: We are going to ask lots of questions like that; we are trying to ask the other one. We are very much trying to ask that.

The Chair: Is the glass half full or the glass half empty?

Alex Sobel: Are there any examples where it can be used positively?

Dr Sarah Kiden: If I use the example we used in our written response, which was using AI for age assessments, the Home Office said that using AI for age assessments would be cheaper than any other option, which is probably true. However, you need to think about many things before you start implementing it. First, if somebody has gone through trauma then the way they age is different than having grown up in a healthy environment. If you use AI in that instance, then you do not know if the system will identify them correctly as underage or over age. Secondly, these systems are trained on large amounts of data. If you do not have the right datasets then you cannot get the results. If you have trained it on sub-Saharan African people and somebody comes from the Indian Ocean region then they might be misidentified or classified correctly.

Alex Sobel: You are basically saying no; Professor Fong, do you have any views?

Professor Kevin Fong: My co-presenter, Dr Aleks Krotoski, would say that the technology is neutral. It is not the technology; it is how we use it. You might as well ask if a machine gun can be used to protect.

Baroness Kennedy of The Shaws: That is the old business, is it not? Gun lobbyists in America say, "It's not the gun that's the problem; it's human beings who use them inappropriately", and use that to justify and resist any constraints on gun ownership. It is slightly worrying that we are going down the same road here by saying that it is not the technology that is the problem but the awful people who put it to bad use.

Alex Sobel: I was going to rephrase it: can AI be trained to protect human rights?

Professor Kevin Fong: I am in no way attempting to defend American gun control; I want to put that on the record.

Alex Sobel: Yes, let us move away from that.

Professor Kevin Fong: What I am trying to say here is that this is the truth about all technology: it is neutral. It is about how we deploy it. Yes, the possibility exists of applying it in a way that it might one day, in some application, protect human rights. I cannot imagine it off the top of my head right now, but that possibly must exist because this is a powerful general purpose technology.

The Chair: If any of you come up with examples and want to write to us about them afterwards, we would be very pleased to hear from you. I am keen that we should hear from my colleague, Lord Dholakia. Then, after that, we are going to hear from Sir Desmond Swayne.

Q22 **Lord Dholakia:** Thank you, Chair. Both my questions relate to healthcare matters and are directed at Professor Fong. Should the use of artificial intelligence in healthcare be subject to sector-specific rules and regulations and, if so, why? Do consumers trust the use of artificial intelligence systems in healthcare, if they are even aware of AI being used?

Professor Kevin Fong: On the first question of whether AI in medicine should have sector-specific regulation—absolutely. We already have sector-specific regulation for medicine. That is because the long history of technology in medicine teaches us that where we do not regulate firmly we get catastrophe at scale. The lessons of the 1950s and early 1960s in the deployment of the drug thalidomide began to give us the current regimes of pharmacovigilance and regulation that we have today. That is partly why medical innovation has to move so gradually: we know that we cannot move at pace in the same way because of the irreversible damage that we potentially can do. That approach should apply in the field of AI when applied to medicine, partly because you end up with irreversible decisions and actions that can cause harm at scale. This technology has the potential for benefits at scale so, on the other side, it also has the potential for harm and injury at scale. Yes, there should be sector-specific regulation. How we get the maximum benefits and reduce the risks is the huge question, especially if we want to try to move quickly, which most people want to do.

Your second question was how we engender trust. That is really an issue of how we educate the wider population and the through-coming generation in the use of these technologies. As with all previous technologies, it is about familiarity and us being able to understand their benefits, risks, limitations and capabilities. You do not do that without improving the population's AI literacy. For that to happen among our patients, it has to happen among the general population in the first place,

so there is a job of work to be done in providing a better battery of education, engagement and familiarity with these technologies in this new era.

Finally, trust is also a product of co-creation. If you have technologies dropped on us from 36,000 feet by the Silicon Valley bros, there will be no trust. These technologies have to be co-created with the people who are closest to the consequences of them—the people who use them and the people they are deployed upon—sometimes with benefit in mind. That is what trust is about: co-creation, building the technologies around the intended target population, and wider AI literacy that itself comes from the education and engagement of the public in general.

The Chair: In case you want to add to that, Mr Birtwistle, the next question is going to be directed to you, so it would be helpful if you could put the two together.

Q23 **Sir Desmond Swayne:** Is it better to regulate in uncertainty or to wait at risk of missing the boat?

Michael Birtwistle: That is classic question of AI regulation. I will come to the previous question at the end. We would say that our understanding of current AI risks is mature. It has been seven years since the Government set up the UK Office for Artificial Intelligence at the then Centre for Data Ethics and Innovation. It has been two years since the Bletchley summit and the AI Safety Institute, now the AI Security Institute, was established. We have had two international AI safety reports. The risks I described earlier have considerable bodies of scientific research and evidence around them, so there is little possible reasonable doubt about their existence and causes. There is a frontier of less well understood and evidenced risks that accompanies the frontier of the technology: extreme risks that could affect critical infrastructure and systemic risks like AI's impact on climate. Some are plausible and potentially high impact if they happen, and it is sensible to anticipate and prepare for them. As a starting point, I do not accept the premise that we do not know enough about AI to regulate.

A second point is about how you regulate. Debate often centres on an unhelpful false dichotomy between letting the impacts of AI run wild or bringing in some imagined heavy rulebook that squashes innovation. There are many ways to regulate and legislate to support that. You can legislate for no regrets requirements that you know you are going to need in any case, like transparency, particularly from the largest companies. You can have framework Bills, like the AV Act which was passed in the last couple of years, that set up the institutions and competencies needed to enforce rules and standards, which you can empower them to write and certify later. The Government can give themselves more regulator powers to remove products from the UK markets that they can hold in reserve without necessarily hitting people with it pre-market.

The last point I would make on this is the sense of urgency about "why now". There is a growing public trust gap between how people feel about AI and the ambition for adoption. There is growing evidence from public polling: our own evidence, nationally representative surveys with the Alan Turing Institute, and polling from the Tony Blair Institute, Labour Digital and the *Guardian* all speak to this trust gap.

In our research, 72% of people said that the number one thing that would increase their comfort with the use of AI is regulation. That is up 10 percentage points in two years, which is incredibly significant in polling terms. When you ask people what they mean by that they say, "Someone other than the AI developers deciding whether AI is safe, powers for Governments or regulators to remove unsafe products, and greater transparency and control".

The Chair: Thank you very much. I was asking whether, in addition to everything that is in the written evidence about polling, we can be sure that we have all that information. It would be very valuable to the committee. Perhaps you can speak to our special adviser about that and we can ensure that it is incorporated.

Q24 **Baroness Kennedy of The Shaws:** I kicked off with that question about layering: there are those who have the algorithms and those who take it into different areas of interest, then there are the people who run the companies that make the profitability and say, "I didn't know about the algorithms and how they might be formulated". It is about who takes responsibility for building in safeguards. It draws on the answer to Sir Desmond's question: you do not want to stymie the development of things that could be very useful for society, yet at the same time, as all the polling suggests, it is felt at every level that some sort of ethical requirements should be built in about what the human rights implications might be, what the damage might be to society, and what the downside in uses could be. Human ingenuity means that not all things will be foreseen, but at least one has people struggling with those things as these particular technologies are being developed. Is that possible?

The Chair: As we are trying to do here today as well.

Baroness Kennedy of The Shaws: Is it possible, with all those layers and stages in the development of AI, that everyone can be educated to be alert to it and to know that there will be obligations which may have implications for them professionally in the long run? It is like reflecting on "Oppenheimer" and the invention of the atomic bomb; it is almost as if that film was made not to look back in time but to look at now. Are we dealing with the ethics of what could happen?

Dr Sarah Kiden: I was going to say that it is possible to assign responsibility at different phases of the life cycle. For example, if you are in charge of procurement then you can require the developer to do certain things. If they do not, then you do not buy the tool from them. You could even tell them that you want to future-proof the system so that if it needs to be used in a different way you can build that into the

system. It is possible to do that, but you need the person at that level to do what they are supposed to do. For example, they have to ensure, at the design and development phase, that they have done enough co-design, worked with participants, gathered the right requirements, and so on. The people who are doing the training on the data also need to train on as diverse data as possible: if it is about demographics then you capture the diverse demographics in the country or whatever it is. At each phase, it is possible to assign responsibility to the person who is in charge. If failure happens, you will know if it has happened at the training level, at the design level, or at retraining or refining. It is possible to do that, but it is not happening right now.

Professor Kevin Fong: To come to your point about lessons from history, the question is: have we seen this movie before? There have been multiple general purpose technologies across recent history: nuclear energy, the internal combustion engine, and the advent of electricity. All required regulation. We are, in this room, surrounded by electricity that adheres to certain regulatory principles, otherwise we are all going to get electrocuted. So you need regulation; the question to ask—and I do not know the answer—is: is this technology sufficiently different that you need to regulate it in a prospective way? It is fundamentally different. The speed of onset and the directness with which it reaches the people who will be impacted by it are different from all other previous technologies. The question is: are the mechanisms of regulation that we have had sufficiently agile and fast to keep up with it? Our previous approach has been to deploy the technology, then the downsides become obvious over a period of time and post hoc regulation comes in.

Baroness Kennedy of The Shaws: Is this not why human rights are so important in these arenas? The questions we are thinking about are: is this detrimental to the human condition? Is this going to damage or create some degrading of humanity? Is this going to affect how people behave towards each other in a way that is detrimental to our respect for them as human beings? Those are the reasons why discussions and the language of human rights provide us with tools. Do you agree with that or not?

Professor Kevin Fong: Wholeheartedly.

The Chair: You agree as well?

Michael Birtwistle: Yes. You can build those mechanisms in but you need to incentivise them. We already do this in a lot of other sectors. We have described some already; planes is another good one. You need obligations on those further up the value chain. MHRA is a good example of a safety case regulator that is able to look at its value chain. Most other regulators cannot. You will probably need to tweak the liability rules to ensure that they incentivise the management of risk post-market as they are supposed to. We are bringing a report out about that in November. Procurement rules are another great mechanism for incentivising, and you need to resource regulators to be responsive. The

EHRC has stated that its funding means that it is limited in its ability to respond to AI issues.

The Chair: Thank you, that is helpful.

Q25 **Afzal Khan:** Let me start with you, Professor Fong, first with a general question and then with a specific question. Are the public concerned about unregulated AI in society?

Professor Kevin Fong: They appear to be. A relatively recent study was done by the Oxford Internet Institute, which surveyed over 3,700 people as a representative slice of the population. Michael and Sarah can speak to it more, but it asked people to identify their political beliefs and whether they self-identified as being left or right of the spectrum. Both sides of the spectrum overwhelmingly agreed that they wanted to see more regulation. Michael has already given the statistics: 72% of all the respondents agreed, whichever side of the political spectrum they identified to being from, that they wanted to see more regulation. This comes back to the trust issue and the general sense of unease the public seem to feel about this technology at this time. They are looking for some reassurance. At the moment, in the absence of anything else, they appear to be looking towards improved regulation or improved regulatory frameworks.

Afzal Khan: Before Dr Sarah and Michael come in, let me give the supplementary questions as well, then you can deal with them as you feel appropriate. Do people feel able to enforce their rights in relation to AI? How can the power imbalance between individuals and big tech be addressed to ensure that rights are respected?

Michael Birtwistle: I will not repeat myself on those statistics, but if you look up "attitudes to AI"—I will make sure it is shared with the committee—then there is a wealth of data on public attitudes to AI. The numbers on things like regulation are very significant. We are talking 88% plus on questions of whether they want regulators to have powers to stop harm rather than private companies. That is an extraordinary majority in polling terms. We are releasing some polling next month that tries to dive a bit more into this question.

To go back to the question of what makes people feel more comfortable, the next answer after regulation is that people want information about decisions made about them by AI. That is very high in public consciousness in terms of what they want to prioritise.

In the research we are bringing out next month, we asked trade-off questions. The public want AI to prioritise fairness, safety and positive social impacts over and beyond economic benefits, speed of innovation, and even international competition. It is always interesting to test what they think when they are trading off. There is a strong sense of public disenfranchisement around AI: many people do not feel they have a say in what the Government do, and it is coupled with high levels of concern about whether the Government will prioritise relationships with big tech.

As to the second question on redress, there is a sense of disenfranchisement if you are talking about public attitude research. If you are talking about legal analysis, we published a legal commission from the AWO Agency about a year and a half ago that found that the mechanisms for exercising your data rights and human rights will not function in most cases for most people. It is very difficult to get redress in practice. It is very expensive and you are unlikely to succeed. The Government, in the Data (Use and Access) Act, have now broadened the set of conditions under which automated decisions can be made about people. That is likely to be a magnifying problem as more and more deployers of technology roll out automated decision-making.

Dr Sarah Kiden: To answer that question, I can use a project we did at Southampton. We used a method called story completion to see what people thought about the future of AI agents. The prompt was, "The year is 2035, you wake up one morning and your agent has done something extraordinary". We were interested in hearing positive as well as negative stories. People were concerned then about the things they are concerned about now, so data protection and responsibility. Some even said that researchers were always running these co-creation activities but, at the end of the day, deployment happened with big tech, so what was the point of participating in co-creation? They felt that they did not really have the power to shape design decisions, even with co-creation exercises, because big tech is making a lot of the decisions.

The Government have the power to require certain things from big tech, and I want to use the example of the USB-C in the EU. If you remember, two years ago the EU said that you had to start selling phones with USB-C, but Apple had always used lightning cables. Apple said it was not going to do that. The EU said that if it wanted to sell in the EU it had to, and the next time it released phones they had USB-C, so you can require big tech to do certain things. If the regulation says they have to do it, and sometimes fines are involved, they might listen.

The Chair: That is a very helpful reply, thank you. Professor Fong, do you want to come back on Mr Khan's questions?

Professor Kevin Fong: Yes, principally to talk about how easily individuals find themselves able to hold these people and businesses to account. These are some of the largest businesses that have existed in the history of the world. In the interviews and contributions that we have had, people are sceptical about their ability to take these people on.

That is borne out by experience. The Norwegian I mentioned earlier in this evidence found it very hard to get redress for what had happened to him. We have spoken to others: we spoke to a woman called Jess Smith, a Paralympian who was born without one of her lower arms. She found that AI was unable to enhance pictures of her without reinserting an arm because, at that time, it could not conceive of an individual with only one arm. When she contacted the AI company responsible for that software it had very little to offer her in return. That shows you a problem of bias in the datasets that were used to train these tools, with an end-user

suffering as a result but having very little redress and opportunity. When you are the size of some of these companies, you can just shrug at these individuals. That is another question: how do you give Governments and regulatory authorities, if that is the way you go, the teeth to engage to give people redress?

The Chair: I know my colleague, Lord Sewell, is going to ask you another question about regulation, so let us go to him now.

Q26 **Lord Sewell of Sanderstead:** How can AI regulation fulfil the goals of balancing its benefits with risks? How do you balance this whole notion of public trust and future risks?

Dr Sarah Kiden: Balancing benefits, risks, safety and trust through regulation is complex. On this panel, we have talked about how AI is an emerging technology. It evolves really fast and there are many complexities. Let us take three different uses of AI: chatbots, self-driving cars and age assessments. These are three different use cases of AI but they all require different things. If you are going to regulate then you cannot regulate the AI used in chatbots in the same way as self-driving cars, because some are more safety critical than others—some we have talked about are catastrophic. The risk level is different for all of them. This may feel like a roundabout response, but the impact of harm is different so there is no standard way to do it.

Lord Sewell of Sanderstead: Self-driving cars are interesting. It is probably just a question of time: they will just be there and we will think nothing of them. Is that a factor: we are not used to it yet and it is simply an instinctive thing? Human risk is about fear balancing benefit. In the end that—and publicity—might be enough.

Dr Sarah Kiden: I would not say that it is about fear. I do not know if you have read about self-driving cars: there are examples where a self-driving car sees a shadow and assumes it to be a human being. In that case, there is not enough training for you to know what the self-driving car will do. It is just a difficult question.

The second thing I want to say is that to measure safety you need evaluation metrics. Up until now, even in the technical community, people have not known what a good metric for safety is. It is not just safety: what do metrics look like for trust, for responsibility, for accountability and so on? Many things are involved in that process.

Finally, I will share an example of a project we did at Southampton with colleagues from Nottingham and King's. We wanted to develop metrics for cultural sensitivity of text-to-image models. We thought we would engage with cultural experts and come up with a list of evaluation metrics for what an output of a text-to-image model looks like. Very quickly, the experts emphasised that you could not measure cultural sensitivity the way you do technically, because in technical terms it is probability, percentage or rankings but when you talk about culture it is more

subjective and qualitative. No communication is happening between the social and technical, in that sense.

Lord Sewell of Sanderstead: Michael, what you said is interesting and kicks in another question about the pharmaceutical industry. If you dare go online and look at the risk elements of a drug then you would not take it. Some of us are taking pills where, if we were to dwell on information, there would be fear. Is there not a sense, in this balance between benefit and regulation, that we probably need to just go out there, take some risks and it will all be well on the night?

Michael Birtwistle: The value of health labelling on medicines comes from a long history and experience of risk management in the medicines and medical devices sector. The reasons why people feel a sense of trust in planes, medicines and food is that they come from safety regimes that try to create a sense of justified trust in those technologies and systems. That is often their objective. You have the trustworthiness element: you know about the risks, you have tried to manage the risks, and you have a professional who is able to prescribe to you despite that risk because it is in your best interests, all things considered. It balances that risk for you. You cannot do that if you do not know about those risks and you do not have the mechanisms to incentivise people to care about them.

The thing that is in everyone's interests here, in terms of keeping up, is having responsive, empowered, well-resourced institutions that are able to give not just the red light to risky stuff but the green light to stuff that is in a grey area. If you look at, for example, what small businesses are currently telling the EU Commission around the simplification agenda in Brussels, which is all about the simplification of the digital rulebook, they are basically saying that what we need is someone to be able to tell you to go faster and say, "Actually, this is okay. This isn't okay". People feel doubt when you have a lack of clarity and overburdened regulators.

A good example in the UK around AI is facial recognition. We had the Bridges case in 2020, the judgment for which said that facial recognition was unlawful in these circumstances and these are the boxes you have to check to make it lawful. It explicitly considered equality and human rights law. Five years later, we are about to have the Government consult on introducing regulation for that technology because police forces have been complaining that the law is so unclear it stops them having the beneficial use of that technology.

Lord Sewell of Sanderstead: The Met recently said that the technology was great at the carnival and stopped a lot of harm.

Michael Birtwistle: That is what the Met will say in public, and it is at the forefront of wanting to use this. If you speak to some of the other 43 police and crime commissioners or forces, many of them are much more nervous—certainly in private—about deploying technologies. They feel like they are being asked to make judgments with considerable human rights implications that should be taken at a national policy level. A lot of the incentive for the Home Office is based on that.

Q27 **The Chair:** You have certainly given us a very interesting overview, and it is very helpful to have the "all right on the night" response to the question Lord Sewell put to you.

In preparing for today's meeting, I took the trouble to go back to what Geoffrey Hinton and Stephen Hawking had said. They talked not about things being all right on the night but about the potential extinction of mankind as a result of AI running ahead of itself in the way they forecast. I do not want you to comment too much on that but I want you to help the committee, because we are now on the final examination question: if you were in our shoes, what would you be recommending to the Government so that it is not just all right on the night and that we guard, as best we can, against the worst possible outcomes?

Professor Kevin Fong: Fundamentally, we need to ask ourselves, "Do we need a different approach because of the uniqueness of this new technology?" We do. It is fast; it has a direct impact on populations at scale and very quickly; it has unique properties that we have not seen before in any other technology, and those principally revolve around its ability to play the role of a convincing human actor, even if it has no human component to it. That said, we have talked mostly about the downsides of the technology but it is clearly going to have potentially huge advantages in the way it may boost economies and fuel productivity and growth in GDP.

How do we strike the balance? If we are suggesting that we would somehow partly abandon the precautionary principle because of the great benefits that we might reap then all we are left with is a containment policy: you must have a mechanism for containment, because things will go wrong and they will go wrong at scale. You then have to ask yourself, "What is that containment mechanism?" My suggestion would be this proposed idea of the sandboxes: of creating a special environment in which Governments and Administrations can learn about this technology and the businesses that are deploying them. At the same time, those businesses can understand, in a safer environment, how to interact with Governments and deploy these technologies. That is probably the best way to have a containment mechanism. I quite like this idea of sandboxes because it gives opportunity. No one knows what is going to happen in the future, so the ability to experiment in some way in what is essentially a safe space and then provide some agility to the whole thing is the right way forward. It gives you a new problem, which is how to construct that safe space, but that is probably the best solution for the moment.

The Chair: Are you giving thought to how you construct those safe space sandboxes?

Professor Kevin Fong: Not me personally, but in reading across the best balance of trying to get the economic advantages as well as mitigating some risks then this concept of a sandbox in which experimenting can take place under close surveillance is probably the

best compromise you can get at the moment. I have no suggestions on precisely what the form of that box should be at this time.

Dr Sarah Kiden: The questions should be: why do we want to use AI in this case? What are the benefits? What are the risks? Look at the AI life cycle and who is involved at each of the phases of the life cycle: where does responsibility lie? How do we know when things go wrong? What does success look like?

Another recommendation would be a stronger link between technologists and policymakers because, looking at the life cycle, it feels like regulation is happening towards the end, which is deployment and maintenance. Maybe this conversation needs to happen way before—at the ideation and design phase—with more investment in public engagement and training people to understand how AI systems work.

Michael Birtwistle: I have two recommendations, if we have to stick to two. The major gap in our regulatory system is managing risk upstream with those building the tech. We need to have a system that incentivises those who are best able to manage the risk to actually do so. That looks like some of the measures I have already described, but pre-market requirements, standards, powers to remove from market and so on. The second—I am going to sound like a broken record—is resourcing and empowering the existing regulatory ecosystem. There has been considerable work inside government, inside regulators and outside government with the Alan Turing Institute on what regulators need in terms of scope changes, upgrades to their powers and the actual resourcing to address the impacts of AI.

The Chair: I thank you all; you have given us terrific evidence. Never mind artificial intelligence; it has certainly stretched our intelligence. It really has been helpful. I know that the committee would be grateful to you if you have any further thoughts that you think might be of use to us as we proceed with our inquiry and get to the point of writing our recommendations to the Government. I thank you all for being here. With that, I conclude this part of our proceedings.