HOUSES OF PARLIAMENT

Joint Committee on Human Rights

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

Wednesday 21 January 2025

3.05 pm

Watch the meeting

Members present: Lord Alton of Liverpool (Chair); Juliet Campbell; Lord Dholakia; Tom Gordon; Baroness Kennedy of The Shaws; Afzal Khan; Baroness Lawrence of Clarendon; Lord Murray of Blidworth; Peter Swallow; Sir Desmond Swayne.

Also present: Dame Chi Onwurah.

Questions 64 – 74

Witness

I: Alexandria Walden, Global Head of Human Rights, Google.

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 witness

Alexandria Walden.

Q64 **Chair:** Welcome back to this afternoon's meeting of the Joint Committee on Human Rights and to our session as part of our inquiry into human rights and the regulation of artificial intelligence. We are privileged to have with us Alexandria Walden, global head of human rights at Google. This is the fourth public session of the AI inquiry that we are staging since the call for evidence closed in September. Today, members will question Google's global head of human rights about what steps the company is taking to ensure that AI is developed and used in a manner that does not adversely affect human rights.

Google, of course, is a household name in the UK and most people are highly familiar with its search engine. Most of us probably use it at least once or twice every single day. However, its AI services, including Gemini, DeepMind and Google web services, are less well known and understood. We would like to explore in this session the themes of transparency, privacy, accountability and equality.

The committee notes the excellent work being undertaken by other Select Committees across both Houses of Parliament in the area of artificial intelligence. For instance, the Foreign Affairs Committee is looking at disinformation diplomacy and how malign actors seek to undermine democracy, and exploring the use of misinformation and disinformation campaigns by state and non-state actors, while the Women and Equalities Committee is looking at the influence of things such as online misogyny. So a whole range of people are looking at various aspects of AI, but obviously our mandate is to look at the effects on human rights.

In that context, I would like to ask you a question to kick off before I turn to my colleague Sir Desmond Swayne to burrow a bit deeper. Does the UK's approach to AI regulation, as outlined in the *AI Opportunities Action Plan*, strike the right balance between supporting innovation and protecting human rights? If you think that is so, are current UK regulations adequate to address novel and emerging forms of AI?

Alexandria Walden: Thank you for the question. If I may, I will take the liberty of backing up and sharing a little about the scope of my role and some of the work that happens as part of our broader human rights work.

Thank you for inviting me to be here to talk about our work on human rights and the intersection of AI. I lead our global human rights policy work. That means that this is what I am thinking about on a regular basis, so I appreciate the opportunity to share it and be in dialogue with you about it.

Baroness Kennedy of The Shaws: Are you a lawyer?

Alexandria Walden: I have a legal background. I went to law school, but I have been in public policy work throughout my career, so that is

what I am focused on. But, yes, having a legal background is useful for understanding all the work I do on a daily basis.

Google has been committed to human rights since its founding. The founders were clear at the outset in their letter to stakeholders that the purpose of the company and our technology was to provide benefit to people around the world. That is true to what the mission of the company is: to organise the world's information and make it universally accessible and useful.

We think that, building on that mission, generative AI will further that aim. Specifically, it can help boost productivity but also enhance expression, creativity and the quality of what we see across the core of services that we provide. We are also particularly excited about the core, cutting-edge work that we are doing on AlphaFold, which is based here in the UK. As you probably well know, the founders of AlphaFold received a Nobel Prize for Chemistry for that work, and the work continues today.

We believe our approach to AI must be bold and responsible, fostering innovation while protecting security, safety, privacy and, of course, human rights. In this work, we are guided by our AI principles. We rolled them out in 2018 and have created updated principles that we rolled out last year. Across those principles, we are focused on ensuring bold innovation, responsible development and collaborative progress. We need to be focused on all those points equally.

The focus of Google's human rights policy team, which I lead, is advancing the company-wide strategy on human rights across all products, including our AI work. We do that through a variety of activity, which includes counselling products on identifying potential human rights impacts, undertaking and overseeing human rights due diligence, and engaging directly with stakeholders in this space.

As Google continues to bring AI products and services to do exactly what we hope—to fuel creativity, expression and productivity—we have established and continue to iterate on an infrastructure, processes and an AI governance structure inside the company to ensure that we have continuous improvement in being bold and responsible. This is true in particular with respect to use by children and our younger users, making sure that we are informed by internal specialists and external experts who have expertise on child development. Our governance structure is really focused on ensuring that we are being bold and responsible and continuing to iterate as the technology develops.

Thank you for giving me an opportunity to lay out the big picture of how we are thinking about it. To your specific question, we agree that the UK's approach on AI and AI regulation is appropriate in taking a context-based approach. As industry, we would recommend this approach as one that achieves this balance between being bold and responsible.

Chair: Is it striking the right balance between protecting human rights and innovation?

Alexandria Walden: We believe so.

Chair: Supplementary to that, if you had to point elsewhere in the world—you have been complimentary about the UK—is anywhere regulating AI in a way that will protect human rights? We know from your background that, before working for Google, you did work on civil liberties and a whole range of things that you can imagine would have great appeal to this committee.

Alexandria Walden: It is hard to point to any specific regulation, but we have been clear from the top of our company that regulating AI is necessary but must be done well. In doing so, we have done a lot of thinking about what we think and would recommend about how any Government or policymakers approach AI regulation. There are a number of things. The first is identifying where there are potential gaps and verifying that there really are gaps in the current regulation, because anything that is illegal without AI is also illegal with AI. That premise is an important starting point. We really do not want duplicative laws or to be reinventing the wheel. That is the core of what we think any regulator or policymaker should be focused on in developing regulations.

Chair: Thank you. Let me go to Sir Desmond Swayne and, after that, I think Baroness Kennedy will want to push you further on the responsibilities that go hand in hand with all the opportunities that corporations like yours have.

Q65 **Sir Desmond Swayne:** If I were to engage or use any service or application, ought I to be able to know whether it was provided in any way with AI? If so, how could that be enforced?

Alexandria Walden: The transparency aspect of what you are asking about is core to the values of Google. We have traditionally been a leader on transparency. One of the foundational pieces in how we think about transparency is that it really matters what the transparency is for, who the audience is and the context in which they are getting it. We have done things such as produce a transparency report on our AI principles and responsible AI which tells, from a broad perspective, how we are creating frameworks and policies, enforcing them and testing around them. That is the big picture.

We also have things such as model cards, which are focused on more of a technical community but create an overview of the model and what it is tested against. Finally, in-product ways to be transparent, specific to the users, are another area where we have continued to iterate and think about really important ways that we can educate users about how AI is showing up in their interaction with our technology. I will give an example of what that looks like, because it is important to ground ourselves in specific applications, which is more effective than trying to say what transparency looks like broadly. The technology is so generally applicable and has broad use cases, so we really have to ground ourselves in a specific use case.

For our consumer-facing products—any way that you might be creating or looking at audiovisual content—we would want you to have information about whether or not that is synthetic. To do that, we have created a tool called SynthID, which allows the user to know whether content is synthetic, created by one of Google's products. That is an important way for a user to look at a piece of content and understand whether it is synthetic. We embed it in our products at the pixelated level, so it is not visible to the human eye, but a user would be able to identify it if an image is showing up in Google Search, for example, or if it is a video on YouTube. Our work on SynthID is an important way to get at users' ability to understand whether content is synthetic.

Q66 **Baroness Kennedy of The Shaws:** You can imagine, Ms Walden, that we have been receiving evidence from all manner of experts and lawyers and people who have some serious knowledge about this. At the end, we will be making recommendations. You are talking about regulation; obviously, we will seek to be as sensitive as possible as to how that might be done. How does Google see its responsibilities when it comes to the impact and outputs of artificial intelligence produced via your own AI services or ones that are shared on your platform? I am particularly thinking about those shared on your platform, because you have described the principles that you have established for Google but, once you have partnerships and associations with other entities, the question remains as to whether they live up to the same standards as you. How do you do that?

Alexandria Walden: There are a few different tools in the toolbox for how that gets addressed and how we think about that. One is that it is important for us to do what we can do across our products, which I began to talk about in relation to SynthID. Another is what we do in concert with other responsible actors and industry. An example of that is our work with C2PA, the Coalition for Content Provenance and Authenticity. We have done that work through a variety of actors, such as the Partnership on AI and others in the NGO community, as well as others in the private sector, to create a protocol to identify and have information and metadata related to real images so that a user can know whether something is a true image, perhaps taken by a journalist or an artist or a photographer. It is about our ability to distinguish between synthetic and real, non-synthetic content.

The creation of that protocol was years in the making among a multi-stakeholder group, to really make sure that we were impacting the broader ecosystem. That is just another example of what we can do on our own, and then what we can do within the ecosystem to try to strengthen the ability of any user to know what they are looking at, whether it is on our platform or on anyone else's. The beauty of the SynthID example that I was sharing is that, if a Google-generated image content video shows up on anyone else's platform, the SynthID watermarking is embedded in that image. If it is circulating on some other platform, any user can look and see whether or not—

Baroness Kennedy of The Shaws: I am going to tell you the honest

truth about what I am digging at. We have just had—you have been, I am sure, very aware of it—publicity around a police force in the Midlands deciding that they have to investigate the conduct of a visiting football team, so the senior police officer passes down the line to some less senior person, "Find out if this football team and their supporters have other instances of really terrible behaviour". Somebody goes off, and we do not know whether they Googled, whether they got your AI folk in or whether it was other some other platform, but inquiries were made and what came up, it turns out, was what is now called a hallucination: a fantasy event which did not exist. I do not know if you know about this, but a football match was described, with terrible events that took place at this football match—and it turned out that the football match had never taken place at all. We got this business of an algorithm that had gone mad. Somewhere down the corruption process, we got information being presented as truth, and it was discovered that it was not true at all.

We have other examples of that with people using AI. I can talk about it in relation to law, where people have investigated things and have based their advices on cases that have never existed—case law that is fantasy rather than reality. We are told that this is basically about, once you have created algorithmic responses to things, they can start developing, if you like, tracks of progress that may not be actually based in reality. There is some serious anxiety here. Who do we sue? Do we sue Google if something like that happens, or do you pass the buck and say that it is the responsibility of somebody else?

Alexandria Walden: Thank you so much for that additional context. There are a few things I want to share in response to that. First, in relation to hallucinating, there are two things. I come back to the SynthID point, because it is important for Google and the responsibilities of others in our sector. We are all iterating here, and this is an example of what others can look to as they think creatively about what they can do. If you see an image, you can upload it into Gemini and ask it whether or not it is—

Baroness Kennedy of The Shaws: The product of AI?

Alexandria Walden: Of Google's AI. If it has a SynthID watermark, we can identify that. That is one way that anyone who receives an image would be able to check whether or not it is synthetic. Now, that is for Google-generated content; obviously, if it is generated by someone else, Gemini would not necessarily be able to identify that for them—SynthID.

I will just get at the hallucinations piece. Additionally, on the user interface of Gemini, it says very clearly and explicitly that the product may make mistakes. It is important to educate users on how they are using the product, so that they are not over-relying on it. For example, for a lawyer drafting a brief through Gemini, it is important to understand that Gemini may be very helpful in summarising generalities about the law but that is not in place of someone who is legally trained—

Baroness Kennedy of The Shaws: Who has expertise.

Alexandria Walden: Yes. Having that explicitly on the user interface is important. Also, we have source links specifically in the prompt response so that it makes clear, whatever summary is being provided, that there is also an external link that you can check to verify and validate whether or not you agree with the summary that you have been provided.

Baroness Kennedy of The Shaws: Who has the duty? I am going to give you another example; I hope that the public who are watching take this on board just now as we come up to the end of the tax year. You give your information to a large firm of accountants and are told that you owe £5,000 and you think, "That is a bit out of the ordinary from what I pay". The person then goes on inquiry or gets somebody else to have a look at it, and finds out that they are not supposed to pay £5,000 at all but in fact some rather inventive young accountant person just did it using AI. If one had paid that money, let me tell you, getting money back from the tax authorities is hell. Who is responsible then? The accountants for using the AI? Or do you say, "Use at your own peril"? Are there going to be warning signs on the things that you make available? How does the consumer—the ordinary person—safely use this stuff?

Alexandria Walden: There are two ways that address a little bit of what you are talking about. One is that, from a regulatory perspective, we think a sectoral approach is best. In so far as we are concerned about tax advisers and how they are using AI in the course of their work, there are regulatory bodies that oversee the activity of that field, and they should be creating guidance and/or anything additional to govern how AI works in that field. That is true certainly in finance as well as in the law, where we have a set of norms, regulations, policies and guidance for the field. How AI gets used in that field should be taken up by that field, because it may look different in the law versus in tax advising or some other field.

Chair: Sorry, a Division has just been called, unfortunately, in the Commons, which means I have to suspend the sitting of the committee while my colleagues in the House of Commons go and cast their votes. We will resume as soon as they return.

Sitting suspended.

Chair: I apologise for the delay, and there are going to be more votes in the House of Commons. For those people joining online, this is one of those things that happens in Parliament. To our witness, I give our apologies for the disruption in the evidence you are giving. We were coming to a final question from Baroness Kennedy before we go to Baroness Lawrence.

Q67 **Baroness Kennedy of The Shaws:** Ms Walden, I was really interested in one of the things that is concerning parents. We know that litigation is beginning now with platforms around the ways in which people can be influenced in the commission of suicide by things that they acquire online. I just wondered: if I were to type into Google "How can I terminate my life?", what would Google's response be?

Alexandria Walden: We have spent a lot of time working on specifically this issue. We have policies in place that prohibit returning information about how to facilitate suicide and self-harm specifically. This is true for all users, not just under-18 users. Obviously, this is not information we would want to be returning for anyone.

Baroness Kennedy of The Shaws: Do you have age verification? If it were, for example, presented in a more subtle way, is there age verification on certain subject matters?

Alexandria Walden: We have a variety of policies that apply for all users, and a specific set that apply—

Baroness Kennedy of The Shaws: For the young?

Alexandria Walden: Yes. Because we take so seriously all of the concerns around ensuring that kids are safe online and that we are adequately addressing the types of harms that are specifically manifesting for kids and teens, we are focused on ensuring that we have an age-appropriate experience for those users. That means that we have done a few things, including rolling out safe search by default. That means that the answers that need to be safe and prohibiting this type of suicide and self-harm content would be true for everyone.

Baroness Kennedy of The Shaws: Adults as well as children.

Alexandria Walden: Yes.

Chair: Thank you very much. Now I call Baroness Lawrence, and after that we are going to hear from Lord Murray.

Q68 **Baroness Lawrence of Clarendon:** Welcome. This is just following on from some of the questions that Baroness Kennedy has been asking you. How does Google mitigate or address unintended bias and discrimination in in AI model training data?

Alexandria Walden: This is an issue that we have spent a long time talking about and take very seriously, so I appreciate that you asked about it. First, because we understand that this is still an emerging transformative technology and it is posing these evolving risks and complexities, we have to have processes and a structure in place to identify these types of risks and address them throughout the product development and deployment lifecycle. That is how we think about it, and we ensure that we are doing it across the lifecycle. We are obviously iterating and learning as the technology is developing. In order to mitigate harms generally and unfair bias specifically, we have a rigorous approach that is focused on both the design-phase testing and monitoring, and then additional safeguards that are focused on ensuring that we are not having unfair bias in the outcomes or results of the work of our models.

Across the model development work that we are doing and product launch as well, we ensure that we have human oversight, due diligence

and feedback mechanisms. Even if things that we tested for were not producing bad or unfair bias during the testing, if challenges evolve later on we have an opportunity to continue to improve or to fix something that has gone wrong.

We want to make sure always that what we are doing is aligning with user goals and more broadly, recognising that we have a social responsibility and that we are aligning with international law and human rights. That is another layer of how we think about ensuring that unfair bias is not permeating our models.

The other thing that I just want to underscore is that we know that the amount, variety and completeness of training data makes a difference in the performance of the models. It is true that, the larger the volume of the training data, that leads to more ability for nuanced and higher-quality results. That is also part of how we think about ensuring that our overall process and the ways that we are implementing it address unfair bias.

Baroness Lawrence of Clarendon: What would happen if an AI decided to go rogue, even though you have the human side looking at it? How do you mitigate that?

Alexandria Walden: It is a good question. That is why we have this big-picture approach to AI governance that is focused on ensuring that we are doing work pre launch in the model development phase, to have executive oversight of what the risks are and how we are mitigating against them before we launch a product. Then we also have processes post launch to ensure that we are continuing to monitor any residual risk and any new and emerging risks, and that we are continuing to learn from that process. Then, because there are executives involved in that process, we are continuing to evolve the infrastructure. Any ways in which we are learning that there are challenges, those are opportunities for us to improve the processes, the policies, the models or our structure overall.

Chair: Thank you very much. I am keen that we have to be swift now because time is against us, unfortunately. We are going to bring in Lord Murray, and then after that Dame Chi Onwurah, who is the Chair of the Science, Innovation and Technology Committee in the House of Commons and is guesting with us today.

Q69 **Lord Murray of Blidworth:** I can reduce my question to this; it builds on some of the answers you gave to the last question. What, in your view, are the top three core human rights risks posed by AI, and what guardrails has Google got in place to respond to them?

Alexandria Walden: I love this question and I will do my best to answer quickly. In general, the way that we think about human-rights-related risk at Google is that we are focused on the pre-existing human rights frameworks, namely the UN Guiding Principles on Business and Human Rights, which govern and provide guidance to companies about how we

should think about protecting and respecting rights. That is the core and the foundation of how we think about this work. When we think about human rights risk assessment, we are grounded in those frameworks as well, so we are focused on the potential negative implications to the enumerated human rights. Also, we are focused on evaluating based on scope, scale and remediability, which is consistent with the UNDP's framework as well.

There are 30 enumerated rights. Because of the breadth of what AI is and the variety of use cases, any given right could be relevant in any given use case. Most often, the salient rights are freedom of expression, privacy and non-discrimination. Again, it really depends on any specific use case. That is where you can get to appropriate mitigations, when you are more focused on specific use cases.

The other thing I will say is that these are broadly questions obviously not specific to Google but relevant for the whole industry. That is why we have engaged with the UN Office of the High Commissioner for Human Rights through the B-Tech Project. One portion of what they do is engaging with companies to make sure that they understand the work that we are doing and how the current UN frameworks are useful in that context. The B-Tech Project put out a foundational paper on generative AI where they discuss exactly this, and there is an annexe paper that spoke of a taxonomy of rights. That is also focused on specifically enumerating rights that may be impacted by AI and providing examples to that. That is just to reinforce that there is really important work being done across the industry, and then specifically at this international organisation, to make sure that we all have visibility into the variety of ways that AI may have an impact on rights.

Lord Murray of Blidworth: Just turning then to the guardrails side of the question, what software guardrails particularly come to mind to you in relation to those top three areas of human rights?

Alexandria Walden: Like I said, those are the three that are the most salient often, but it could be a variety. In particular related to those three, though, the best way that we address it is through that structure that I talked about where we have model requirements, and that is where we are ensuring that we are vetting and filtering quality data and that we are testing the models and doing evaluations based on safety, security and other harms, including harms to rights.

Lord Murray of Blidworth: What are the limitations of those types of guardrails, in your view?

Alexandria Walden: For Google, we have an infrastructure in place to ensure that we are only launching products where we have had an ability to assess the risks, appropriately mitigate those and ensure that we are monitoring them in an ongoing way to address anything that comes up post launch. Really, we require an infrastructure to be able to have that process in place so that we are preventing those harms.

Lord Murray of Blidworth: I am aware that there are ISOs in relation to AI. Are there any other universal rules that would assist in the risk management of the human rights risk posed by AI?

Alexandria Walden: The primary places are the UN, the OHCHR—

Chair: Order, order. We are now inquorate.

Sitting suspended.

Chair: Order, order. I am calling the committee back to order, because we now have two colleagues from the House of Commons who have raced back from the Division Lobby, and we are very grateful to them for that. I shall use my discretion in the chair to move around some of my colleagues' questions. They are all important, but we have to try to get through as many of them as we can before we complete the business. Mr Afzal Khan is here and has a question to you about the protection of children.

Lord Murray of Blidworth: Sorry, Chair, I think that the witness was just about to finish her answer.

Chair: Okay. Let us keep it very brief though please, Lord Murray—but thank you. As you can see, this situation is not of my making.

Alexandria Walden: There is the UN Office of the High Commissioner for Human Rights, and there is also work happening at the OECD and the responsible business workstreams there—and then there is the work happening at the standard-setting bodies. The UK has been there for much of this and has helped to lead it, and there is ongoing work in that regard, but there is also a focus on human rights there too. Those are the core places where we are seeing efforts to focus on the pre-existing international standards and ensure that they are relevant to this conversation.

Chair: Thank you for that, Lord Murray.

Q70 **Afzal Khan:** A way of protecting children from adverse outcomes to their fundamental human rights is to restrict access to products and services that use AI. In what circumstances should AI systems and products be subject to age restrictions? If your answer is yes, at what age should restrictions be set and how could restricted services be operationalised effectively?

Alexandria Walden: Thanks for that question. For Google, our goal is that we are providing age-appropriate experiences for our users, for kids in particular. What that means is that we have put in place specific product features or guardrails that are working to address those harms and ensure that kids have a safe experience online. We have done that through a variety of things such as parental controls, implementing safety by design, providing enhanced privacy protection and, lastly, doing things like promoting AI literacy among children. They are growing up in this digital age, so ensuring that they have the opportunity to understand

how to use these tools to their benefit and to navigate them is really important.

On that front, we have a programme called Be Internet Awesome; we have evolved that work for the AI age, and we have a new programme called Be Internet Legends, which is focused on kids and minors using AI and teaching them how to do that responsibly. We have partnered with ParentZone here in the UK to launch that to 10 million kids, so it has reached a large percentage of users. Again, ultimately ensuring that kids understand how to use and navigate the age-appropriate services is the best way in which to address these issues.

Afzal Khan*:* If restrictions are placed on products and services that use AI, what are the potential net impacts for human rights?

Alexandria Walden: It is a good question, and it is one that we have only begun to have across the wider human rights ecosystem. The key piece is really that we need to be focused on the human rights of a child, which means that they have freedom of expression rights and privacy rights. We need to make sure that when we create regulations and restrictions around the use of this technology it still allows for them to engage, learn and discover as appropriate for their age. All the ways in which adults can benefit, kids can benefit too—we just need to make sure that we have age-appropriate restrictions.

Chair: Thank you. As a supplementary to Mr Khan's important question about children, I was shocked to see this piece appear in one of our national newspapers—that Google had groomed children on turning off parental controls. Have you seen that?

Alexandria Walden: I have not seen that article.

Chair: Perhaps we can give it to you and ask if you would be good enough to write to the committee afterwards in response. It may be more misinformation—I do not know—but it is a serious question and we would like to hear the answer.

Alexandria Walden: We would be very happy to help.

Q71 **Dame Chi Onwurah:** Thank you, Ms Walden, for giving evidence to the committee today. Not all AI providers are as ready to hold themselves accountable as Google seems to be, so I want to thank you for that. I have noted your reluctance to give answers based on principles as opposed to specific use cases. My question is with regard to LLMs generally. Given Google's mission, as you set it out, to organise the world's information, and the specific use case of adolescents and vulnerable people who are turning to LLMs for counselling and advice, if I ask an LLM a question, will it tell me the truth?

Alexandria Walden: I appreciate the question. This gets at the crux of how we think about how we build these tools to create and increase access to information and fuel productivity while also being responsible in addressing this concern. So there are a few things here: if you ask a

model a question, it will return an answer. The reason why that is important is that it gives us an opportunity to ensure that we are educating the user on exactly the experience that they are going to be having.

Dame Chi Onwurah: My question was about the answer and the nature of the answer, and whether it would be truthful. You have already made points about education and other points. Would the answer be truthful? Would it be the truth?

Alexandria Walden: We absolutely strive to have fact-based answers, but it is also true that the models may hallucinate. From a transparency perspective, it is important for us to be clear explicitly on the face of the product that that may happen so that the user ensures that they have the ability to source tech any answer that is being returned.

Dame Chi Onwurah: I accept that Google puts a little get-out clause at the bottom of its service to say that it may not always be the truth, but will the answer make it clear that that particular answer may not be the truth? How can people trust what that answer is?

Alexandria Walden: Our goal is to provide high-quality responsive answers that are fact-based to our users. For example, if we are looking at AI overviews, which are grounded in search, that is where the answers are being pulled from. That is why we have the source links within the prompt responses, so that it can be clear for you to fact-check or see what the reference is that the model is pulling from. We absolutely seek to provide quality answers.

Dame Chi Onwurah: I say truth, you say quality.

Chair: It is often said that the English and Americans are divided by a foreign language—but I think we both know what Dame Chi is getting at.

Dame Chi Onwurah: Perhaps you could write to us with your understanding of truth.

Alexandria Walden: I would be happy to follow up.

Chair: We would be very interested to see that. Our colleague Dr Swallow wants to ask about large language models, which is the point that we were just talking about, so it might be advantageous to bring him in now, and then Mr Gordon.

Q72 **Peter Swallow:** The reason why this is such an important question is that we have already seen that it can have significant real-world consequences. Last week, we saw the West Midlands Police chief constable stepping down after it emerged that his force had relied on evidence from an AI that was false—which was an hallucination. I should emphasise that the LLM in question was produced by one of your competitors, not by Google. Nevertheless, it throws into stark relief the real-world consequences that AI hallucinations could, can or do have. If individuals are relying on incorrect, fabricated or manipulative outputs of

foundation models, who should be held responsible for that?

Alexandria Walden: Thank you for providing some additional context on the question. There are a few things here, or two things. One is the overall governance structure that we use to ensure that we do testing, both pre launch, identifying any risks prior to launch and addressing them, and then continuing to monitor them after the fact. That is how we improve the models and ensure that quality is improving over time—that there is less hallucination and fewer errors. The process that we have in place, the infrastructure for testing, gets at that.

The other piece is building tools to help users, whether they are journalists, prosecutors or your average person, understand whether an image is synthetic or real and whether it is AI-altered. That is some of the work of the SynthID tool and the C2PA protocol.

Peter Swallow: Just to be clear, we are not talking about images necessarily—we are talking about information. It is as simple as the fact that it used to be the case that when you put a search into Google you would get a list of sources that you could then click through and have a look at, making your own deductions. It is now the case that the first thing you see in very many cases is Google's AI summary of that information. That is fundamentally different, is it not? Does it place a different responsibility on Google and other companies to recognise their role in making sure that information is as accurate as possible?

Alexandria Walden: From our view, it is important for there to be responsibility apportioned, but that should be differentiated between different actors in the ecosystem. When for example someone shares a piece of content that would otherwise violate our acceptable use or our guardrails, it seems unreasonable that the developer would be responsible there, or even that the deployer would be, if the person who created the prompts and then disseminated it is the one who circulated the information. Ultimately, there needs to be appropriate responsibility for different actors across the ecosystem based on what the harm is and who is more culpable.

Peter Swallow: So if companies' guardrails are circumventable, that is not the responsibility of the company.

Baroness Kennedy of The Shaws: Why should not those who make profits be the people who are held responsible?

Chair: Sorry, we are really against the clock. If we have additional questions on some of these things, I am sure that you will be happy to correspond on that.

Alexandria Walden: I would absolutely be happy to follow up.

Baroness Kennedy of The Shaws: But that is the key thing—why should not the people who make profit be the people who are held responsible? They would then make sure that the products being created reach the right kind of standards. It is about profitability—follow the

money. Why should not the company be held responsible?

Alexandria Walden: We believe that the responsibility looks different for different actors across the ecosystem. Then it is about really defining what the harm is, which helps us to get to the point of where responsibility should lie. So depending on how we are defining what the harm is, we can focus on who is most responsible—and responsibility should look different for different actors, depending on their relation to that harm, as it is defined.

Q73 **Tom Gordon:** In 2023, the current CEO of DeepMind signed a statement from the Centre for AI Safety which warned about the possible risks of extinction from AI. That is something that, as an MP, I have had many emails from constituents campaigning on that issue. What should we be saying to those people? How are you at Google managing existential risks to human rights such as human extinction? To build on what we have just heard, if Google or other foundational model developers are not responsible, who should be?

Alexandria Walden: The way that we think about this is twofold. First, we have infrastructure in place to do AI governance, so that we are focused on what the risks are, ensuring that we are identifying them, testing against them, monitoring and litigating. That is from the pre launch to the post-launch process. That is the core piece of how we should think about risk. Those are the risks that are before us today, and part of that is thinking about foreseeable risk. There are longer-term risks that we are also focused on. There are two things to point the committee to—first, a paper that we did that focused on artificial general intelligence and our framework approach to that, as well as some additional papers that we have published relating to our frontier model framework for how we think about identifying and mitigating severe risk. We are certainly thinking about those bigger picture and longer-term risks as well, while we also have an infrastructure in place to ensure that we address the harms that we are concerned about today.

Chair: That is very helpful. I think we would like to write to you about superintelligence to try to probe some of that further, because it is something that the committee is troubled by.

Two of my colleagues had questions particularly about opt-outs for individuals to express that they do not want to be subjected to AI. A colleague had a question about how you manage the differing rules for AI across different jurisdictions, and how Google deals with that. Perhaps if we could, my colleagues would be happy to correspond with you on that, as time is against us. I turn to the last question for today.

Q74 **Sir Desmond Swayne:** Could you make two recommendations for our report?

Alexandria Walden: Thank you so much for the opportunity to share on this. It is not often as a human rights practitioner we get this opportunity, so I really do appreciate it sincerely. First, we think that the UK is taking the right approach; we think that identifying a way in which

to have a pro-innovation regulation focused on being sector-specific and context-based is the way to go—so we reinforce that. The other piece is that, as a human rights practitioner, I would say that you should continue to ensure that you are explicitly baking in respect for human rights into language. Using the UN guiding principles on human rights as the baseline framework for that is important, because that is how companies that are focused and ready on human rights and responsibility understand how to implement that across their business. Certainly, there is space for additional companies to take that on as well.

Chair: Thank you very much indeed. You will hear the bells ringing again—it is not one of the prisoners trying to escape; it means that our House of Commons colleagues again have to go and vote. They are working pretty hard this afternoon—I hope viewers will take note. Thank you so much for being with us today, Ms Walden. It has been a real pleasure to hear from you. We will correspond, if we may, on those other issues. With those words, I end the session with the words "Order, order".