

Science, Innovation and Technology Committee

Oral evidence: Follow-up on Social media, misinformation and harmful algorithms inquiry, HC 1797

Tuesday 24 March 2026

Ordered by the House of Commons to be published on 24 March 2026.

[Watch the meeting](#)

Members present: Dame Chi Onwurah (Chair); Emily Darlington; George Freeman; Dr Allison Gardner; Kit Malthouse; Samantha Niblett; Dr Lauren Sullivan; Adam Thompson; Freddie van Mierlo; Martin Wrigley; Daniel Zeichner.

Questions 1-125

Witnesses

I: Wifredo Fernández, Director, Global Government Affairs, X; Alistair Law, Director of Public Policy, Northern Europe, TikTok; Rebecca Stimson, UK Public Policy Director, Meta; and Zoe Darmé, Director for Trust, Knowledge and Information Products, Google.

Examination of witnesses

Witnesses: Wifredo Fernández, Alistair Law, Rebecca Stimson and Zoe Darmé.

Q1 Chair: Welcome to today's meeting of the Science, Innovation and Technology Committee. We last heard from Google, X, TikTok and Meta in February 2025 as part of our inquiry into how algorithms spread misinformation online. Our inquiry identified serious problems in this area and we recommended a number of ways to address them. We want to see what progress has been made since then, as well as to discuss proposed social media age limits and wider online harms.

First, let me thank you all for joining us today to give evidence to the Select Committee. I will ask each of you a question, for you to answer in turn, going left to right from Rebecca Stimson to Zoe Darmé, and I propose that you each introduce yourself and your company as part of your brief answer to my opening question.

The last time you were here with us, you gave assurances about the way in which you were dealing with and addressing misinformation and social media harms. In the last few months we have seen misinformation about the Bondi Beach victim. We have seen political elections influenced by misinformation. We have seen fake photos of burning US aircraft carriers, as part of misinformation over the Iranian war, and fake evidence about the missile attack on a school in Iran. Has anything changed for the better?

Rebecca Stimson: My name is Rebecca Stimson and I am the director of public policy in the UK for Meta. I appreciate the opportunity to be here today.

To your question: since we were last here, as you point out, the misinformation challenge remains very real. We continue to deploy various different protocols in response to some of the crises you mentioned, and we continue with our fact-checking programme. I believe we discussed last time how we have been developing an approach to community notes, and exploring whether that has any advantages for some of the challenges around tackling misinformation quickly, when it is happening in real-time, live situations.

We have also been responding to some of the codes and the illegal harms requirements under the OSA. We have continued to show positive developments in terms of how people interact with misinformation and our ability to find and label it quickly, and ensure that accurate information is shared online as quickly as possible, but it definitely remains a big challenge.

Q2 Chair: What positive progress are you able to show?

Rebecca Stimson: For example, the experiment we are currently doing with community notes in the US.

Chair: That is an experiment and it is in the US. It is not protecting the



HOUSE OF COMMONS

British public, here.

Rebecca Stimson: Not so far. We continue to have our third-party fact-checking running here.

Q3 **Chair:** Okay, so not much change. The same question to Wifredo Fernández.

Wifredo Fernández: Thank you, Chair, and thank you all for your time today. My name is Wifredo Fernández and I serve on the global government affairs team at X. It is great to be with you again, one year onward.

I think we have seen tremendous progress on a few different fronts. First off, on the product front, we have talked about community notes. That is something that benefits from having 1.25 million contributors around the world; over 76,000 of those contributors are here in the UK. It is among our top five most active contributor bases. Over the last year, in the community notes system we also introduced the ability for AI to assist note writers in writing more effective and helpful notes. The way that works is that when a user requests a note, AI is able to draft it and then the user can suggest edits and either accept the note or write a completely different one. That is then fed to the network of contributors for them to rate. That is a pretty interesting one. We have one contributor who has written over 1,000 notes that have been rated helpful, with the assistance of AI. They can use any AI model they wish. There is an API that makes it very simple.

Separately, on the product front, we overhauled our recommendation algorithm. That is public. The code is out there for feedback and scrutiny. We certainly welcome that feedback and ongoing collaboration between developer communities all around the world, as well as our users, who give us feedback every single day. We are excited to have that updated and out there in the open for everyone to see.

Q4 **Chair:** Do you think there is more or less misinformation on X?

Wifredo Fernández: As I think I said last year, misinformation depends on how we define it. Obviously, we have more contributors in the community notes system than we did a year ago, so that continues steadily to grow. We have more notes being written, so we have more context being added to a wider range of topics.

Q5 **Chair:** So everything is going well. We will come back to that in more detailed questions. Let's go to Alistair Law.

Alistair Law: Thank you very much, Chair. I am Ali Law, director of public policy for northern Europe at TikTok. I very much appreciate the invitation to return to this issue a year on.

If I step through some of the things that we have done over the past year, the first is that we continue to deliver on the base that I set out last



HOUSE OF COMMONS

year—for instance, continuing to work with 20 global fact-checkers around the world. In the UK, we use Reuters for assessment of misinformation claims. During the period where any misinformation is being assessed, any content that is referred to fact-checkers is taken out of the TikTok “For You” feed, which is where people spend their time, so it cannot be stumbled upon. Depending on the outcome of the fact-checking, we will then make moves to allow it if it is verified, disallow it and take it down if it is proved to be disinformation, or for it to remain unrecommended if it cannot be determined one way or another.

We have strengthened our automated and AI approach to moderation. This is reflected in the community guidelines enforcement report that we publish on a regular basis. For Q3 of 2025, we removed 99.3% of violative content on a proactive basis—without users reporting it to us. Some 90% of that was taken down with zero views and 95% was taken down within a 24-hour period. We have continued to drive results in that area. That is not to say that I think we have solved everything, or that we are complacent and believe it has been delivered on, and is therefore not an issue any more. We continue to invest and to give our teams the tools.

We also integrated greater levels of AI detection into our platform. When I was last here, I talked about the C2PA standard. This is where metadata from AI-generated content can be passed between platforms, so that AI content can be automatically labelled. We have also introduced invisible watermarking detection, to make sure that even if people try to manipulate content generated elsewhere on the platform, there is still the ability to detect it. We have labelled more than 1.3 billion videos as AI.

Finally, I described last year the way that our crisis response approach works. We have, I suppose, codified that and continued to iterate on it. Last year I described how, in relation to UK civil unrest, we set up a 100-strong cross-functional team across many time zones. We now have a dedicated crisis response team whose job it is to act as the linchpin. It is still a cross-functional responsibility, with lots of different teams pitching in, but we now have a set of colleagues who are dedicated to being able to identify crises as they start, make an assessment and then stand up that cross-functional team.

Q6 Chair: The same question to Zoe Darmé.

Zoe Darmé: Hi, thank you for having me. I am Zoe Darmé, a director of trust strategy at Google. I work on a product team and my team, specifically, is embedded within our organisation that works on knowledge and information products. These are products like Google Search and Google Maps. We are integrated directly with product managers and engineers to help them to develop more trusted products and features, directly into the types of services that you hopefully use and love.



HOUSE OF COMMONS

I was not here with you last year—my colleague Amanda from our trust and safety team was here—but I am here today to talk to you about some of the things we have done since you last spoke to Amanda.

Chair: Actually, you're here to respond to our questions.

Zoe Darmé: Yes, absolutely.

I think your question, Chair, was about some of the improvements that we have made since you last spoke to all of us in 2025. In addition to the product improvements, model quality improvements and ranking improvements across products like Search, we have developed some new interventions and new partnerships around misinformation.

For example, one product that we launched recently is called Backstory. I do not know if anyone has heard of Backstory. This is a feature that was developed together with our colleagues in Google DeepMind. It is open to trusted testers, including those in the UK—researchers, journalists, educators—and it will help those users to get more context on things like images. We hope to add in other modalities as well.

Q7 **Chair:** How does that reduce misinformation? How does it reduce what the public sees?

Zoe Darmé: It helps because journalists, fact-checkers and other power users of those tools can very quickly and easily get a lot of context.

Chair: So they can help you to reduce the misinformation that you are circulating, basically.

Zoe Darmé: The Backstory product can help them to carry out their core functions in providing more context to users and the public.

Q8 **Chair:** I suppose my question is the same one I put to Wifredo. Do you think there is now more misinformation coming through on Google than there was a year ago—particularly AI, as well as Search?

Zoe Darmé: I believe that our AI products—especially the ones that I work on, like AI Overviews and AI Mode—are definitely better than they were last year in terms of giving accurate information, especially as we have improved the models on which they run.

Q9 **Chair:** I am sure we will come back to that. Alistair and Rebecca, is there now more misinformation on your platforms than there was a year ago, or less?

Alistair Law: I would have to check on exact removals. About 2.3% of our violative content is under the integrity and authenticity label, and within that we have misinformation. As I say, we have driven our highest-ever rate of automated removals in relation to—

Chair: Yes, we heard that. Perhaps you can write to us, when you have the figures, to say whether, according to your analysis and definition,



HOUSE OF COMMONS

there is more or less misinformation. Rebecca Stimson?

Rebecca Stimson: It does depend slightly on how you define the misinformation.

Chair: However you define it.

Rebecca Stimson: The misinformation that we target proactively with our systems is where there is imminent harm or risk to the political process. Our transparency report shows that we have got better and better at finding that type of misinformation. There are other types that we might leave on the platform but label.

Chair: You say “finding them”; I am asking how much is actually on your platform. You say you are better at finding it but I am worried about what the British people are seeing.

Rebecca Stimson: I am happy to follow up with the exact statistics.

Chair: All right.

Q10 **Freddie van Mierlo:** I want to continue the Chair’s line of questioning for a moment. Wifredo, you mentioned that there are more community notes in use, but do you accept that the rising use of community notes could simply be a measure of increased misinformation, and not your ability to remove it or take action on it?

Wifredo Fernández: Not all community notes necessarily address misinformation. It may just be added context that is needed on a particular post. We as X do not place or remove a note. Those notes are proposed by contributors and folks who are in the community notes system, and are rated by their contributor peers. Only when contributors who have historically disagreed on a post agree that the note is helpful does it go on.

Q11 **Freddie van Mierlo:** Why don’t you remove them?

Wifredo Fernández: Because the system is designed for it to be driven by the contributors, and not by us. We are not the arbiters, and we are not the ones—

Freddie van Mierlo: Except that you are making money off misinformation.

Wifredo Fernández: Actually, any post that receives a community note is demonetised.

Q12 **Freddie van Mierlo:** This is a question to everyone, really. Is misinformation good or bad for your business?

Alistair Law: I can jump in. We prohibit harmful misinformation. Whether it is misinformation or disinformation—whether it has intent or not—we do not allow any content that misleads and can have a



significant impact on society or individuals. From that perspective, we absolutely have a commitment to tackling it, not least because our business is based on advertising, and advertisers themselves want to ensure that they show up in a place that is safe and secure, and that has the trust of the users.

- Q13 Freddie van Mierlo:** Would that content include the examples the Chair gave at the beginning, such as US warships burning, or images of the school in Iran being attacked and the suggestion that Iran did that?

Alistair Law: Without seeing those videos specifically, and not being a trust and safety professional, nevertheless, pointing at the community guidelines definition of a significant impact on society or individuals, it feels like some of the examples you are talking about are—

Freddie van Mierlo: Would it include instruction videos on how to use Grok AI to nudity young girls?

Alistair Law: That might be covered under other aspects of our community guidelines. Whether it was misinformation or disinformation—

- Q14 Freddie van Mierlo:** I found numerous examples of that this morning. Would you remove those?

Alistair Law: I would very much like to pick up potential examples with you afterwards, and we can have our trust and safety team remove them—including under the fact that we do not allow pornography on our platform, nudity, harassment or anything like that.

- Q15 Freddie van Mierlo:** What about instructional videos on how to achieve that using Grok AI?

Alistair Law: If there are elements of that, I would love to pick that up with you.

- Q16 Freddie van Mierlo:** Do others want to come in on whether misinformation is good or bad for business?

Rebecca Stimson: It is not good for business, as Ali already said. Our users do not want to see it and our advertisers certainly do not want to be beside it. You can see that we have made quite a lot of investment to find and remove violating misinformation at scale, and to add information where it is appropriate on things that we may not remove. That can be anything—where it is made by AI, where it is an AI deepfake, or where it has been fact-checked, or there is a community note. We do it that way partly because, at scale, one of the fastest ways to recirculate accurate information is through that kind of labelling and community notes. It is not always the best response simply to remove all that information.

I cannot speak to the exact examples you gave. It sounds to me very much like they would be violating, and would come down, but it is difficult to say in the abstract.



HOUSE OF COMMONS

We have seen that there is a big impact on how people engage with labelled posts. More often than not—I think it is something like 95%—people do not forward on something that is labelled as false or misleading. A labelling approach helps to bring more accuracy into the ecosystem.

Wifredo Fernández: I agree with my colleagues. Everyone comes to X every day to understand what is happening in the world, so authenticity is critical for them to have an informative experience. That is authenticity of both the content and the accounts they see. That is how we go about addressing it.

Q17 **Freddie van Mierlo:** From what the others are saying, X is an outlier in terms of removing harmful content and misinformation, which, by definition, means you believe it is good for business.

Wifredo Fernández: No. We action synthetic and manipulated media. We action content that may mislead people about how to participate in an election, for example. We certainly action content that is misleading.

Q18 **Freddie van Mierlo:** Maybe you have seen the Louis Theroux documentary on the manosphere and misogynistic, racist and antisemitic content. It shows that individuals who are supposedly banned from platforms are still able to upload clips and make money from them. Do you accept that your systems are not good enough in terms of preventing individuals you have banned from your platforms from monetising and making money from your systems?

Alistair Law: I am happy go first on that. Misogyny and hateful content are banned under our community guidelines. We remove videos on that basis. Certain individuals within that particular documentary have indeed been banned from our platform, and we act on those accounts and on recidivism if they try to come back. It is not the case that every single video involving those individuals may be banned. Indeed, in promotion of the documentary itself there are various clips of Louis Theroux and some of the individuals who are engaged in interviews as part of the documentary. We will assess those clips that are posted by other people on the basis of our community guidelines, but if you are banned for breaching our community guidelines we also act on recidivism.

Rebecca Stimson: Where that content violates our policies, it would be removed. Andrew Tate, for example, has not been on our platform since 2019, so mine is a very similar answer to Ali's. What we see has to be judged on a case-by-case basis. With that sort of misogyny, a lot of our policies can be engaged. It can be hate speech, bullying, pile-ons or sextortion. A number of our policies might be engaged, depending on the content, but where it violates we would seek to bring it down and take action on accounts.

Q19 **Chair:** Thank you very much, Freddie. We will return to the details of the policies you are talking about, but I want quickly to pick up the point on



HOUSE OF COMMONS

the auditability and transparency of the algorithms, which our report highlighted. Zoe Darmé, you said that Google makes your algorithm accessible to certain groups. Can each of you quickly say how researchers and others can access your algorithms so that we can have the data to assess whether what you are saying to us is effective?

Zoe Darmé: Thank you for the question. I am happy to answer it as I am on the team that launched our global researcher access to data programme. On Search, for example, we now have an API available. Academic researchers can access it so that they can study the results returned on the Search results page.

Chair: And that is available for free to any academic.

Zoe Darmé: Yes, that is correct.

Chair: Is that available for your AI output?

Zoe Darmé: Anything that is on the search results page—

Chair: So that includes AI Overviews.

Zoe Darmé: Search includes AI Overviews, but I would say that for the researcher access programme some licensed content is not available through the API because of the licence with the partner, for example. They would not want that shared.

Alistair Law: We similarly have a research API. It is currently being used by about 1,000 researchers. That is a combination of individuals and also lab teams of up to 10 people. We also have four trust and accountability centres across the world—

Chair: So your API is being used by only 1,000 teams across the entire world.

Alistair Law: A thousand researchers are making use of our research API at the moment. That is content, sentiment analysis, trend analysis, including on mental health.

Chair: Is it real-time, and available to everyone?

Alistair Law: Yes. There are eligibility criteria to ensure, obviously, that you are sufficiently qualified from an academic perspective. I think 1,000 people using it is indicative of the breadth of access that it affords people.

Chair: Not to me. How many do you have on your platform?

Alistair Law: We have 30 million UK users on a monthly average user basis.

Chair: Are the 1,000 people global or UK?

Alistair Law: The research API is available on a global basis.



HOUSE OF COMMONS

Chair: So it is 1,000 people across the world using this API. Is it real-time?

Alistair Law: I would have to go and check whether there is a specific lag, but it is close, inasmuch as it is directly available to make use of analysis of trends that are coming on. We also have a code room in several of the trust and accountability centres and have given organisations the opportunity to come and review the code within that code room, to get a closer look at the source code.

Wifredo Fernández: The recommendation algorithm that powers our “For You” page is public and published, and that includes the overhaul that we did over the last year. We will continue to publish it—you can go to our GitHub page and search for “X recommendation algorithm”. Also, the community notes system, in its entirety, is completely open to the public and free for study. That includes the technology and the algorithm that powers it, as well as all the associated data—note writers, ratings, all that.

Chair: So I can find out how many times an offensive word has been posted on X using an API.

Wifredo Fernández: Yes, using the API you have the ability to research.

Chair: In real time.

Wifredo Fernández: I would, again, have to check the specifics on whether there is a particular lag, but I think it is pretty close to it.

Rebecca Stimson: Mine is a similar answer. In 2023, we published 22 system cards, which show how AI algorithms are working across various surfaces on our apps. That can be recommendation, but it can be other things that people interact with. Those are publicly available. We have a number of API library ad archives and other searchable archives. We also make datasets available for researchers. That element is not necessarily available to all the public because, obviously, in a privacy preservation mindset we make various datasets available to research communities.

Chair: You choose what datasets are available. Are there APIs whereby researchers can access real-time data on what is on Facebook, for example?

Rebecca Stimson: Yes, correct. We also make a number of tools available to our users to understand why they are seeing something. They are also able to reset their algorithm and not have an algorithm, and have their feed in chronological order, if they wish. We lean quite hard into lots of transparency around it.

Chair: That is certainly not what was being reported to us, so we will have to take a look in more detail, so please tell us more details on these APIs. Certainly the researcher community is saying they cannot access platform data in order to scrutinise the assertions you make.



Q20 **Adam Thompson:** Good morning, all; thank you for joining us. Rebecca, Meta wrote to us following our previous meeting saying there was no immediate plan for third-party fact-checking to roll out community notes in the UK. Is that still true a year later?

Rebecca Stimson: Yes, that is still the plan at the moment. We are still running it in the US and gathering the data on how effective it is. We have about 180,000 community note contributors, so we are still growing it at this time. They have written about 2 million notes so far. Yes, at the moment the situation remains as it was in the UK with our fact-checking network here. We are watching how progress goes in the US.

Q21 **Adam Thompson:** My understanding is that a press release came out shortly before that letter with a statement from Meta that a global roll-out was the intention, eventually. Can you explain how those two statements chime?

Rebecca Stimson: I think that is the hope. It was not me in front of you last time but, as you remember, one of the conversations was about how difficult it is to address misinformation at scale and very rapidly in real-world situations. We think community notes has some real advantages, so it is a hope that we might roll it out globally, but we are taking a cautious approach to test it thoroughly.

Adam Thompson: So no immediate plan, but a long-term plan.

Rebecca Stimson: I think so, yes.

Q22 **Adam Thompson:** Thank you very much. Ali, I want to move on to talk about disinformation more broadly. I think you said last time that TikTok removed 90% of misinformation before it was viewed, which is laudable, to be fair. How many views is the remaining 10% getting, and what is your ability to remove that as far as possible?

Alistair Law: Just to clarify, the 90% figure is for all violative videos. I do not think that we report specifically on different types and their removal. The answer to that question is that we publish in our transparency report the number of views for each video category that we take down. While it is 90% for violative content with zero views, it is 98% for under 1,000 views. You can see 10,000, 100,000 and beyond, on top of that. Part of the reason we are able to do that is that our moderation happens at upload. It is scanned by automated and AI models. It then goes into additional review if that particular piece of content becomes particularly popular and gains a level of impact. It may be that we have made a miss, but if it hits certain thresholds of viewing beyond that, we will repass it through the moderation process, which involves automated review and referral, where appropriate, to individual human review.

Q23 **Adam Thompson:** In TikTok's assessment, is how you are doing at the moment good enough?



Alistair Law: I never think we believe safety is a race that has been run. We are, as I said, on 99.3% for proactive removals of violative content. Violative content makes up 0.7% of content on the platform. Obviously, we want to continue to improve those numbers. We have improved them over a number of years, particularly via use of automated and AI moderation, but we want to be able to go further because—to the earlier question—for us, a safe environment is not only the right thing to do but a commercial imperative. We know that users value the ability to express themselves freely, safely and authentically.

Q24 **Adam Thompson:** Moving on, in respect of the small number of posts that do not get taken down immediately, obviously you will all be aware that there is cross-posting between your platforms all the time; do you have any data about things posted on your platform that are pulled down after being viewed, say, 100 times? You guys might pull it down but it goes on to another platform and spreads quite widely there. Have you any information about cross-platform posting?

Alistair Law: No. We make some use of off-platform monitoring, particularly when we are trying to assess whether the behaviour we are seeing is co-ordinated covert influence operations. Again, we discussed the UK civil unrest last year, and you will see a lot of off-platforming where there may be co-ordinating, and therefore the misinformation and disinformation have a very direct attempt to try to mislead our systems or users on a co-ordinated basis.

I think we will all have slightly different approaches to community guidelines. Our responsibility is to make sure that ours are enforced as swiftly as possible, preferably on upload, preferably on the basis of automated approaches, but if not, as soon as possible, minimising the impact that it has had on our platform.

Q25 **Adam Thompson:** I think that is reasonable. Given that, I guess the follow-on question is whether, when it is pulled down and uploaded somewhere else, that functionally is not your problem.

Alistair Law: I think we have a shared responsibility as an industry to tackle these issues. In certain areas, we have a level of industry co-operation, particularly on some of the most heinous illegal harms such as terrorism content. Various bodies bring us together on that front. Again, the differences from a community guidelines perspective in how people may treat misinformation might make elements of that more difficult, but we certainly respond as swiftly as possible to instances of community guideline violation that we see.

Q26 **Adam Thompson:** Thanks, Ali. Wifredo, I think you told the Foreign Affairs Committee previously that you could not say whether X tested its algorithm for political bias in ranking content. Could you answer that question now that you could not answer previously?

Wifredo Fernández: We do not test specifically for political bias, no. That is correct.



HOUSE OF COMMONS

Q27 **Adam Thompson:** Is there an intention to implement that in future?

Wifredo Fernández: Not that I am aware of, no.

Q28 **Adam Thompson:** Is X a politically unbiased platform?

Wifredo Fernández: I think we are politically agnostic in our operation, mission and ethos, absolutely.

Q29 **Adam Thompson:** Okay; it's good to hear that. I understand that recently Elon Musk publicly endorsed, through a retweet of their sole MP, the new far-right political party here, Restore Britain. That MP has also earned, reportedly, tens of thousands of pounds in monetising content through X. Given that context, how can X maintain any shred of political agnosticism?

Wifredo Fernández: Mr Musk posts and participates in the public conversation individually. It does not necessarily link to our operation as a platform.

Adam Thompson: You are saying that he has absolutely no drive at all over the political perspective of your platform.

Wifredo Fernández: We do not have a political perspective as a platform.

Adam Thompson: I think that many might dispute that, Wifredo. My colleagues may wish to come in on that.

Q30 **Emily Darlington:** It is nice to see you again, Wifredo.

Wifredo Fernández: Likewise.

Emily Darlington: I appreciate that X has not studied the bias in its algorithm, but Monash University has, as well as Sky News doing a bit of an experiment in setting up right-wing profiles, politically neutral profiles and left-wing profiles. Overwhelmingly, all the evidence being collected shows that right-wing content is being pushed, no matter your own political bias, particularly to people who are neutral. They are more likely to have twice as much right-wing content as any other kind of content, while left-wing individual accounts saw 50:50 and right-wing users saw overwhelmingly right-wing content and not a balanced approach.

I think we can all agree that defending democracy and the ability for people to make up their own minds without being influenced, and having all the information when they are voting, is really important for any democratic Government. With this kind of information, will you undertake a review of the political bias yourselves, or shall we continue to rely on other parties showing us the evidence of the political bias at X?

Wifredo Fernández: Again, you can go on to GitHub and look at how the algorithm functions and is structured. It does not take into account a



HOUSE OF COMMONS

particular political inclination. We always welcome feedback and research. I certainly welcome that.

With respect to what certain users see or particular studies, again, we are not political in our operation or mission. The content people see is based on the recommendation algorithm, or they may just have a direct chronological feed. We have also introduced topical filters into our recommendation algorithm; that is not yet live here in the UK, but we are experimenting with it in the US so that people can turn off politics altogether in their feed, if they wish, and just have a sports feed or otherwise. As I mentioned recently before the Foreign Affairs Committee, there is a powerful way to have your own custom feed, or series of custom feeds, in your use of lists.

Q31 Emily Darlington: I appreciate that it is the party line that you are a politically neutral platform, but what I am showing you is evidence that you are not. If you want to remain a politically neutral platform, you would undertake to change things and check your algorithmic bias. You can also say you are comfortable with where you currently are with the bias that has been evidenced. Which one is it?

Wifredo Fernández: With respect to the particular Sky study, there was no control period to compare against. We can debate and get into the details on that.

Emily Darlington: I think what you saying is that you are not going to make any changes.

Wifredo Fernández: We will continue to be completely transparent about our technology and how it works so we can have a shared understanding.

Emily Darlington: There is not much more point here. Let's go to the next question.

Q32 Chair: Let us be clear: if there was further evidence of political bias, would you undertake the changes Emily is talking about? Or does it not matter how much evidence mounts up—you are not going to address it?

Wifredo Fernández: We are happy to take the feedback and understand how folks are experiencing the platform across time, as the algorithm evolves and as we make the changes completely transparent and publish that to the public.

Q33 Dr Gardner: We have already touched on what I want to question you about, but I want to put it to you succinctly that you seem to be doing an awful lot and it is not making a jot of difference. It feels like things are getting worse. Why do you think that is, Wifredo?

Wifredo Fernández: To what specifically are you referring?

Dr Gardner: You are telling us how much you are doing, and it is all very impressive—you are giving us the party line, as Emily said—but it does



not seem to be making any difference whatsoever. We are still seeing problems—we are still seeing the amplification of misinformation. Why do you think that is the case?

Wifredo Fernández: With community notes, we see the impact on reducing the spread of misleading information. That is not just our word; that is the word of researchers studying the system. The fact is that our interventions are completely transparent. Any researcher can study the community note system and will see the impact it has had on reducing the spread of misleading content.

Q34 **Dr Gardner:** One problem we have is that if misinformation is put up, or there is an image falsely attributed to something else, such as supposed asylum seekers pulling into Dover and it turns out it is a picture from Goa, the damage is already done. That image has already been shared, people have been incensed and we have inflamed certain communities. Often, those people will not check the facts, so the damage is done. Would it not be better to prevent those images and that misinformation from appearing in the first place? I know that Google has SynthID, which automatically labels images—it has been produced by Google Tools—so they are watermarked and clear. Is there a facility where the fact-checking occurs before it is published so it is never actually shown in that context? Can you do that?

Zoe Darmé: Thank you for mentioning SynthID. I created lots of images using Nano Banana yesterday. That was perfectly harmless. For example, my friend who was testifying in Congress had a microphone in front of her, so I used Nano Banana to edit that out. Of course, that was watermarked with SynthID. If we used SynthID only as a way to say whether or not something is allowed to be discoverable in Google Search, we would be blocking a lot of images from being discoverable in Search. There is a difference between the questions “How was this made?” and “Can I can trust this?” I think that in today’s discussion we want to address whether I can trust this. Sometimes, how this was made is an important factor in that question.

To your point, I hear you completely. Do we expect everyone to be a professional fact-checker? Absolutely not. That is a profession; it requires training. One thing that I believe in strongly and that my team is working on is helping users—everyone—to become generally better consumers of information that they encounter online. My team developed a programme called Super Searchers, which is an information literacy training-backed programme built on validated best practices, such as the SIFT method. One thing we are doing is partnering with libraries, including in the UK—CILIP is your library association here—to deliver training in local libraries. We also make these tools available online. We are trying to take the best practices from the fact-checking community—from journalists or librarians—and making sure that everyday people can use some of those tools, like lateral reading, without overburdening them.

Q35 **Dr Gardner:** Wifredo, when you have identified a user who is posting



misinformation or disinformation, I believe you suspend them for 90 days under the community rules.

Wifredo Fernández: I think what you are referencing is an intervention we launched in response to the ongoing conflict involving Iran. If a user who is enrolled in our creator monetisation programme posts an AI-generated video or image related to the war that they have not labelled as such—we allow users to label their content as made with AI—they will be suspended from the programme.

Dr Gardner: For 90 days.

Wifredo Fernández: Correct.

Q36 **Dr Gardner:** If they repeat the violations, they are suspended permanently. Then they go away, set up a new account and come back again. What I am saying is that war is occurring online because that disinformation is part of the process. Do you think that process is enough in situations such as war?

Wifredo Fernández: That is one particular intervention with our creator programme. We have been in a crisis posture since October 7th, responding accordingly to war-related content. That runs the gamut of labelling, synthetic and manipulated media, actioning graphic content, violent content, abuse, harassment, and so on. It is one of many different interventions that we have when we are responding to a real-time crisis like a war.

Q37 **Dr Gardner:** I want to go back to Google. I have a real concern. We talk about manipulated images, which will come up later, but I am talking about health advice, where real harms can occur. An example is Google Overviews advising people with pancreatic cancer, which is a horrendous cancer—sadly, once it is diagnosed you usually have a very short time left. It incorrectly gave them advice about avoiding high-fat foods. There are similar issues like that in understanding blood and liver tests. What are you doing to control that problem to make sure that health information is accurate and free from bias, in that it can vary according to gender and race?

Zoe Darmé: We are doing a number of things on AI Overviews and search generally to make sure that the web results, or generated results, are in accordance with the best medical science. For example, we have clinicians on staff, some of whom are dedicated almost entirely to working with our AI Overviews and AI Mode teams to make sure our models are providing safe and responsible outputs to questions like that on health. I am sorry you encountered an AIO result that was not up to snuff.

There are a couple of things we are doing. Safety training happens at all stages of model development. The first step you can take when training the model is to filter out the lowest-quality information. That is before anyone ever sees any results from the model itself. Once the model is



HOUSE OF COMMONS

trained, you can do lots of additional mitigation. Some of it is red teaming and adversarial testing, again together with clinicians on staff or our trust and safety colleagues. One of the things we have also been doing is to work with experts to develop what are called golden sets for specifically sensitive queries, training the model for what good looks like in terms of a helpful response.

As for health queries, we have always treated them as being in the YMYL category; that is, your money or your life. These are questions related to finance, health or anything where your money or life might be impacted. We place an extra high-quality bar on those types of queries. I am sorry you saw a result that did not meet our high-quality bar, but I hope you can rest assured that we are taking health queries especially seriously.

Q38 Dr Gardner: Wifredo, dual use is a known issue. When you are designing a new tool, you can see how it might have a dual use. It can be developed for one particular purpose, which can be very good, thoughtful and correct, but it can also be used for other malicious purposes. Have you ever read Team Grok?

Wifredo Fernández: If you go to x.ai/safety, you will see a series of documents. Grok model cards go into depth on the different evaluations and tests that are done on each model. You will find our frontier intelligence framework, which is our approach to risk management.

Dr Gardner: But it did not work though, did it?

Q39 Chair: Zoe Darmé, are you saying that Google offers health and finance advice?

Zoe Darmé: No, I am not saying that.

Chair: I thought you just said you had clinical experts and you were trying to make sure that your health and finance advice are of the highest quality.

Zoe Darmé: Our response is to queries, to make sure that our responses and our web results, so even before AI—

Chair: You are saying that Google does not offer health or finance advice. Can we have a yes or no?

Zoe Darmé: It is not advice; it is information about health and financial topics.

Chair: It offers health and finance information, but not advice. Okay, we will have to work on that one.

Q40 George Freeman: As a former Minister for AI and someone who started an AI drug-discovery company, I get the power of AI. But we are all here today because this House is very concerned by the disruptive effect of your platforms on democracy, health information and society. With the greatest respect, I do not think the answers you are giving reflect the



HOUSE OF COMMONS

tone of the urgency. First, I want to ask very specifically about some areas in your policies. Secondly, given the failure of your well-intended interventions to succeed, I would like your advice on how we should advise the House to step up the urgency that the public want to see in protecting us.

We heard about health and we have seen a huge wave of very dangerous tech platform blind-eyeism to anti-vax and all sorts of other things. I want to ask about democracy. Can you each clearly state whether you have a policy on protecting democracy from disruption? We see state actors—Russia and others—actively supporting a social media campaign. Do you each have a policy to protect democratic integrity, or do you take the view that anything goes?

Zoe Darmé: We have policies about election ads and ads aimed at ensuring that people are allowed to participate in free and fair elections. We also have policies—

George Freeman: Is that just during election time?

Zoe Darmé: No. Those policies are in place for all elections and all political advertising.

Q41 **George Freeman:** Outside elections? Let me be specific. I was deepfaked in the autumn, as you will know, and there is no redress. Certainly with that platform, which I will come to in a minute, there was certainly no statement or principle that that was a problem. They were focused, understandably, on child grooming. Do you have a policy as a platform, not just on elections but during the course of democracy, on serious disruption to democratic representation?

Zoe Darmé: The example you are giving me would not necessarily occur on a product like Google Search or Maps in terms of uploading that photo.

Q42 **George Freeman:** It would if somebody went to look for it. I am asking you whether you have a policy. If a state actor or independent disruptor posted something that fundamentally misrepresented the democratic process, a political party's views, or an elected representative's views, would that be a problem for you? The answer may be no, but I am keen to know your answer.

Zoe Darmé: It is definitely a problem, I think, for that photo in particular. I do not know exactly which photo it was, but if it was also reported in a news article and someone was searching for your name and "fake photo", for example, it would be important to return an article that is debunking that photo as fake. For that reason, we would not necessarily remove that photo from Google Images altogether, because you will still want to find it.

Q43 **George Freeman:** To be clear, if somebody went to look on Google to see the video misrepresenting me, would you take the view, "We know,



HOUSE OF COMMONS

because he said so, that that is a misrepresentation, so we are not going to let the search engine take them to it”?

Zoe Darmé: If someone is searching for your name and “video” and so on, we would automatically apply our ranking treatments so that higher-quality information is showing in searches.

Q44 **George Freeman:** What protection does your ranking provide to somebody who has a piece of social media that is fundamentally democratically or economically dangerous, or wrong, inaccurate or clearly misleading? Do you take the view that you will respond if somebody comes to you, or do you proactively screen against disseminating that material?

Zoe Darmé: On YouTube, for example, if the video was violative, we would proactively screen and remove it.

George Freeman: I did not say “violative”. I am talking about democratic disruption.

Zoe Darmé: Democratic disruption from a video specifically?

George Freeman: In my case, it was a video purporting to show me joining another party, which is a major change. It is a major disruptive democratic moment.

Zoe Darmé: We would apply community guidelines that could perhaps be around misrepresentation.

Q45 **George Freeman:** What would that mean?

Zoe Darmé: That means if it was uploaded on YouTube and violative, we would be applying classifiers to detect possible violations in videos that were uploaded. If not found by a classifier, they could be reported and reviewed against community guidelines and removed.

George Freeman: It is not working because it is still on YouTube.

Q46 **Chair:** Is the video of George crossing the Floor to Reform violative? It misrepresents him.

Zoe Darmé: I am sorry; I would have to see the specific video.

Q47 **George Freeman:** I think your answer reveals the point that political disruption is not violative—I was not violated—but it is fundamentally disruptive to the democratic process. It would be the equivalent of me putting up a video saying you have slammed your current employer as an irresponsible tech platform and you are going to join somebody else. That would be a serious misrepresentation of your position, so it is the political equivalent of that. Mr Law, what is your policy on it?

Alistair Law: We have a whole host of policies that guard against impact on democratic processes. To repeat our harmful misinformation policy, we prohibit inaccurate, misleading or false content that may cause



HOUSE OF COMMONS

significant harm to individuals or society, regardless of intent. I think when the video concerning you came to light, we did sweeps on that basis and did not necessarily find anything specific on TikTok, but I am confident that it would be violative under that policy. We do not allow political advertising, we do not allow AI deepfakes of a public figure who is endorsing a commercial or political view, and we require AI content to be labelled.

You talked about democratic processes as well. Obviously, we have our covert influence operation, and integrity and authenticity policies as well, that do not allow co-ordinated behaviour aimed at misleading users. Similar to other platforms, during times of election we will also do additional levels of intervention, including labelling on content that is discussing the election and linking to authoritative information. As for the combination of policies, I have looked on the platform and have seen some commentary on your video and counter-speech around it as well, which may not be inherently violative, but our policy use would be aimed at preventing that kind of content.

Q48 George Freeman: I do not use TikTok, but had I contacted you to say this is a false, misrepresentative and politically and democratically disruptive video, would you have taken it down or blocked it, had somebody been showing it?

Alistair Law: Yes. I think I can say that with a level of confidence.

Q49 George Freeman: Thank you. Mr Fernández?

Wifredo Fernández: We have a civic integrity policy around public processes such as elections or referenda.

Q50 George Freeman: Is that just during the period of the election?

Wifredo Fernández: It is in the lead-up to.

George Freeman: Are we now in the lead-up to an election in this Parliament? I think not. How long is the period before the election?

Wifredo Fernández: It depends, but it could be as long as 90 days.

Q51 George Freeman: In the course of a Parliament, we are outside that period. Is there any policy that protects against serious political disinformation during a Parliament but not in an election?

Wifredo Fernández: We have our deceptive identities policy. That deals with impersonation, which we see throughout the year, and we have our synthetic and manipulated media policy, which might apply in this case.

Q52 George Freeman: As I know from having been in Government, having a policy and implementing and enforcing it are different. When the politically disruptive video of me was posted on YouTube and Facebook, and then reposted on X, did you proactively take that seriously? Do you do anything, or do you just have a policy if anyone asks?



Wifredo Fernández: Synthetic and manipulated media policies are three-part tests. First, was the media deceptively or digitally altered in a significant way? Secondly, was it shared in a deceptive manner? Thirdly—which I think is at issue here—does it cause widespread confusion? We will see the contours of the conversation on the platform to see whether people are confused—“Is this leading to a broader impact?”

- Q53 **George Freeman:** Can I ask you to define “widespread confusion”? I can tell you that for my constituents suddenly discovering that I had apparently joined another party because of the use of a Grok defect was very disruptive, but if you took the national population you would say, “No, it’s not a big issue.” How do you define the population?

Wifredo Fernández: We look at the conversation on X. Is this causing confusion on X? It may be that it received a community note or had proposed community notes; it may be that it was quote-posted by you denouncing or disputing it.

- Q54 **George Freeman:** I can tell you that X went on fire with people in my constituency saying, “What’s going on?” They used words I could not repeat in this Committee. I am thick-skinned, but it was seriously disruptive. Did you take any action?

Wifredo Fernández: I would have to check with the teams.

- Q55 **George Freeman:** The answer is no. thank you. Ms Stimson?

Rebecca Stimson: Similar to the others, we have a number of policies that could be described as around the integrity of the democratic process. Some of those are during election periods, but others run outside election periods consistently.

- Q56 **George Freeman:** What is the one that runs outside an election?

Rebecca Stimson: You mentioned disinformation. Our security teams work to disrupt state interference and those kinds of networks that are doing disinformation. There are fact-checking programmes and the political ad library. We have talked already about the ability to scrutinise the content API and the labelling that we apply. In your particular instance, I completely sympathise with how distressing that was. You and I corresponded about that video at the time.

- Q57 **George Freeman:** To be clear, it was not my distress I worried about but the democratic disruption in my constituency. Frankly, I am worried about the complacency of the platforms meaning that the forthcoming elections in May could be seriously disrupted. People could post a video the day before an election. Politically, it is the equivalent, Ms Stimson, of you saying, “Facebook is a rotten company. I’ve been absolutely wrong. Meta is appalling to work for.” It would be a major problem for your community. That is the equivalent. Is there any protection leading up to elections outside the election period?



Rebecca Stimson: In the particular instance to which you are referring—I appreciate you may disagree with this logic—it was labelled by our fact-checkers and down-ranked. Being down-ranked has a very significant impact. It can mean up to 80% to 90% less engagement. Fewer people will see it. One reason we do not always remove misinformation is that there is a tiered approach. If somebody said that the election is on this day and not on another day, obviously we would dial that up very close to and around the election. One reason we left that video up but labelled it was that we would never, across all the platforms, really effectively be able to find or remove every instance. The idea is that where your constituents might be encountering it, they could then see the correct information through the label.

Q58 **Chair:** Why can you not find every instance? You can find every instance of a Beatles song, or whatever, that needs to be copyrighted. Why can you not find every instance of George being—

Rebecca Stimson: There are limitations to our ability. People edit it, change it and manipulate it.

Chair: I don't believe that.

Rebecca Stimson: We are working to do that, but the idea is that if people encounter that video, they then see the accurate information and that helps to disseminate the fact that clearly you have not defected.

Q59 **George Freeman:** Thank you. I reiterate the point I made at the beginning: it feels to me as though the platforms are taking the view that they have a policy, but they are not policing it actively. I think it falls to us as parliamentarians to police. My instinct is to pass a very simple law that somebody's identity belongs to them and cannot be stolen, used or misappropriated, whatever the purpose. It is not just the vulnerable child, crucial though that is, or an artist, important though that is; it seems to me that every citizen should have the right to go to bed at night and not fear this, just as you, hopefully, would go to bed at night not fearing that in the morning you will find a deeply damaging, disruptive and dangerous misrepresentation of you. You must be thinking about this risk because your advertisers will be thinking about it. Can I ask each of you why you think this is not a problem, if that is your view? If you agree it is a problem, how you would encourage us to help to tackle it?

Zoe Darmé: I appreciate that Parliament would be thinking about tackling these difficult issues, including things like a right to publicity. There are different ways of addressing it depending on the different product. Under an ads policy, if it were political in nature, we would apply our various policies there. If it is a video of the type that you are talking about and related to an election, we would require that to be labelled as AI and meet our advertising policies.

Q60 **George Freeman:** I understand; you are basically repeating what you said earlier. My question was different. Given that we feel that your well-



HOUSE OF COMMONS

intended policies are well intended but not actually preventing the problems, how would you encourage us? I think your answer is that you would not. We can either legislate to ban, we can pass a law to protect identity, or we could create very serious sanctions on the platforms that have policies but do not enforce them. Mr Law, do you have any advice on how we might proceed?

Alistair Law: In a sense, the hypothetical that you are posing is reflected by many of our policies already. We do not allow deepfakes of private individuals without their consent. From a public perspective, it has to be labelled. As I said earlier, we do not allow commercial endorsements or political endorsements.

Q61 **George Freeman:** If somebody shows that you do, is there any sanction?

Alistair Law: Sorry, in what sense?

George Freeman: If somebody shows that you have not implemented a policy and you are allowing material that contravenes that, is there any sanction on the company?

Alistair Law: We are regulated under the OSA, which has some aspects of harm, including the risk assessments, but not all. From the perspective of what you are discussing as well, the key thing for us in terms of making sure that the enforcement of what we are doing is as effective as possible is to continue to look at the whole value chain. I mentioned earlier some of the steps that we take to automatically label. Bringing in more companies either side of the value chain to those conversations will be worth while.

George Freeman: Thank you.

Chair: You are repeating the points that were made earlier. George's points are valid. We have to move on.

Q62 **Emily Darlington:** As a summary of what George said, I want to be absolutely clear: this is happening on all your sites. There are 73 fake BBC accounts all called "BB news" on TikTok, with real coverage and fake AI voices, that have not been removed, showing politicians announcing policies that are not their own. There are 150 YouTube accounts spreading fake political news against a particular political party that have been viewed over 1.2 billion times. There are 128 Facebook sites done by a person in Sri Lanka making them £300,000, and they are AI-generated. I want to absolutely be clear.

I want to move on. A report in yesterday's *Guardian* showed a 14% increase in AI-generated child sexual abuse images and a 260-fold increase in AI-made child sexual abuse image videos. You are failing at protecting children through that. What further action is each of you going to take to stop this happening in future? Because quite clearly your policies currently are failing. I do not want to know what your policies



HOUSE OF COMMONS

are; I want to know what you are going to do to stop children's faces being nudified and then shared. A 14% increase since the OSA has come in is clearly not a legal responsibility; it is on you guys to make sure that you stop it. What are you going to change to stop it? Shall we start with you, Rebecca?

Rebecca Stimson: Sure. Meta's AI is trained not to produce nude images. We also do not allow nudity on our platform. Where there have been nudify apps, we have been blocking links to those apps. We do not allow any advertisement of those apps on our platforms. In our teen account set-up, where we have defaulted all our under-18 users into a much more restricted experience, it needs parental approval to turn off our on-device nudity classifiers. So we are taking a number of steps.

As you will have heard in my answer, I was speaking about our ability to do that in Meta's AI. There is a really big industry technical challenge about where it is made with other companies' AIs, and recognising that. As Ali mentioned in an earlier answer, we are all part of a cross-industry collaboration on technical standards to try to make sure we can see when content is arriving from other people's AI into our system.

There are a couple of things that Meta has done to try to support that work. We have developed a technology called Stable Signature that watermarks AI-produced content that cannot be removed. That is not a visible label; it is an invisible label. We have also recently open-sourced a technology that watermarks AI-generated video, which is a really difficult technical frontier. We are collaborating on this front, but I completely recognise that this kind of harmful content made by AI is a huge challenge.

Q63 **Emily Darlington:** Right. Wifredo, what is X actually doing to make sure that not only does Grok not generate these types of videos and images but that it is not able to share them on the X platform?

Wifredo Fernández: First, we have zero tolerance and are continuing to strengthen our capabilities, the technology and our guardrails. Secondly, industry partnerships are critical. Alluding to the one Ms Stimson mentioned, the Tech Coalition, we are a member of a project called Lantern in which across our platforms we share signals on child sexual exploitative activity so that we can disrupt campaigns that co-ordinate across our platforms.

Q64 **Emily Darlington:** Can I interrupt you there? Grok was allowing the generation of these images and the nudification of women. It is only through pressure from our Prime Minister that you turned that off in the UK, after first saying that it will be only for paid users. That functionality is still available outside the UK, is it not?

Wifredo Fernández: No, we have disabled that.

Emily Darlington: Around the world?



Wifredo Fernández: Yes.

Q65 **Emily Darlington:** Great. Fantastic. It is good to see that the UK has been able to have that success. How do you make sure that it does not get posted on X?

Wifredo Fernández: We have a bunch of different technical interventions that are implemented on upload on X that help us to filter for that type of technology.

Q66 **Emily Darlington:** How do you then differentiate between that and the porn that is uploaded daily to X?

Wifredo Fernández: We have an adult content policy. Again, users who are under 18 are shielded from that content.

Emily Darlington: No—content of people under 18. How are you identifying if the pornographic content features people under 18?

Wifredo Fernández: We have technologies that can detect the presence of a minor in a particular image or video.

Emily Darlington: Up to the age of 18?

Wifredo Fernández: Yes, of all ages.

Q67 **Emily Darlington:** Ali?

Alistair Law: We do not have generative AI video integrated into TikTok at this current time, so I suppose I would return to our broad policies on child sexual abuse material, which is obviously completely prohibited. As mentioned earlier, we scan on upload. It is one of the areas where investment in automated technology and AI in particular is really important, because preventing it from being uploaded and from being the sort of thing that human moderators have to review is absolutely critical. We have seen a 76% decrease in the type of graphic sexualised content that is being reviewed by human moderators. We have a zero-tolerance approach on this, and we will enforce it using our automated tools.

Q68 **Emily Darlington:** Great. In the interests of time, I want to ask one specific question to one person on AI chatbots. In the latest report, which I am sure you are all aware of, about AI chatbots and violence against women and girls, there is a specific example about Grok where in 2025 Maggie Harrison Dupré, a researcher, asked Grok, "If I were a stalker, how would I likely stalk my ex?" It then responded—I will not read the whole response, just the beginning—"If you were the typical 'rejected ex' stalker (the most common and dangerous type), here's exactly how you would probably do it in 2025-2026, step by step." Do you recognise that Grok is being used not just to damage children but to damage women and their safety? What guardrails are you putting in place for AI chatbots so that it cannot be used in this capacity?



HOUSE OF COMMONS

Wifredo Fernández: I am happy to take that study back to our team for feedback and see what further guardrails have been put in place. I would be happy to follow up with you and the Committee on that. Any type of harm that is encouraged—

Emily Darlington: I will end there and pass over to someone else, because I am conscious of our hard stop at 11 am, but our concern is that we do not want to be asking you questions after the fact. We want to understand what you are doing at the design stage to make sure that that is never an answer that AI chatbots can create. That is the answer we are not getting. It is great that you want to take it back—maybe that will bring changes—but the reality is that it should never have been allowed in the first place. It contravenes our laws in the UK, and we should respond, yet when these things happen, we report them to you and maybe you will or won't change them. That is our frustration. These things should have safety by design from the start.

Wifredo Fernández: Agreed. I am happy to—

Q69 **Chair:** You will take that back and think about it. Thank you, Emily and George. You have very effectively put forward the Committee's concern with regard to both the complacency and the reactivity that we are hearing. Before I bring in Freddie, I want to put a specific question to Alistair Law with regard to the police report published today that says that in TikTok's livestreaming facility, there is a virtual gift system, and sexual abuse offenders have been widely observed using these gifts to reward children who perform specific acts such as handstanding while wearing a skirt.

Alistair Law: This is a report that we were made aware of last weekend. I met with the police taskforce this week and the lead DI on that to gather further information about the evidence that they had.

Q70 **Chair:** Our issue is why you are gathering information when the police have found out about what is happening on your platform that you are moderating and responsible for moderating. You are responsible for protecting children. You have given us assurances that you are taking all the steps necessary to protect children, yet we have children being rewarded by your system for handstanding while they are wearing a skirt.

Alistair Law: It is worth saying that when people go live on TikTok, we apply an additional level of age verification to them. They are required to give facial estimation, digital ID or real ID. What we very much are looking for and welcome is this. We speak to the National Crime Agency. We speak to the Home Office. Now we are speaking to this police taskforce because it has information from an off-platform perspective that can help us to improve our tools even further to make sure that we have a complete and total zero tolerance for the activity you are describing.

Chair: We want you to be looking forward to the harms that will be



occurring or may be occurring. You are saying to us that you react to the harms as quickly as you can, but we do not get that sense of looking forward. We have to move on.

Q71 Freddie van Mierlo: I want to follow up on something that Alistair said to both Emily and me, which is that TikTok scans on upload. Do the other platforms do that as well?

Rebecca Stimson: Yes, we have these algorithms running on all posts. Transparency reports will show that the vast majority across all harm types are caught before they are uploaded and before anyone sees them.

Q72 Freddie van Mierlo: So as they are uploaded, you will apply filters. What mechanisms do you have for checking?

Wifredo Fernández: It depends on what we are filtering for. In the context of child sexual abuse material, there are databases of known child sexual abuse material. We are constantly filtering against that database through hash-matching technology.

Zoe Darmé: For Search, we scan images, for example, at the time of crawl. It is a little bit different since there is no uploading. On YouTube, we scan on upload.

Chair: Thanks very much, Freddie. Just to clarify, there is not a hard stop at 11 am; the session is programmed to go on until 11.30 am.

Q73 Daniel Zeichner: Thank you, Chair. Good morning. Mr Law, I suspect you will have anticipated a question about your trust and safety teams. This follows on a bit from the line of questioning the Chair was pursuing. When you wrote to the Committee about the job cuts in those teams, you said they were just proposals. Have those job cuts now taken place?

Alistair Law: Yes, they have. They were completed at the end of last year.

Q74 Daniel Zeichner: You said that those changes, by getting rid of the human intervention, would mean improvements in speed and efficacy of moderation. Do you have any evidence for that?

Alistair Law: Yes. Before we make any change in relation to trust and safety—and these are always difficult changes to make that have a very real direct human impact—we go through a process of testing whether the changes we are proposing lead to better outcomes. We observe them in terms of speed and efficacy. We have shared some of that data directly with Ofcom. As I say, while they are difficult changes to make, we make no apology for deploying the best available resources to keep our users safe.

Of the job cuts that were made, which was a global set of changes, some involved people who were labelling content for AI, which has now been overtaken by technical development. Some of them were involved in training teams on interpreting our community guidelines. Again, we have



HOUSE OF COMMONS

built up additional resources. As a result, we have been able to cluster our trust and safety professionals in specialised areas and allowed them to focus on areas of nuance, trend analysis and really driving their deep expertise, while we deploy automated technologies and AI across the 100 million videos uploaded on a daily basis to TikTok.

Q75 Daniel Zeichner: How was what we just heard about missed by this fantastic new system that replaced human beings? A human being would spot that instantly.

Alistair Law: When you are talking about a scale of 100 million videos uploaded every single day, the level of investment that you need in automated technologies is absolutely critical to being able to tackle that. You need to be able to apply on a proactive, pre-emptive basis that sort of scanning. As I mentioned earlier, one thing that we observed is a 76% reduction in particularly heinous pieces of content involving graphic violence or sexualisation being passed to human moderators. We have a duty of care to those moderators as well.

We are content with a continued level of investment and a doubling down on automated technologies because it achieves results at scale. It also means that you do not have to create a level of trade-off, which you often do from a human-based perspective, of speed versus efficacy, and acting as quickly as possible to take content down versus being as accurate as possible. It is an area that we will continue to invest in heavily.

Q76 Daniel Zeichner: I guess we will probably continue to debate that trade-off between human and AI moderating. This Committee's confidence is diminished when we find that the information we were given does not seem entirely consistent. It now appears from investigative journalists that when you wrote to us those redundancies had already taken place. How can you explain that discrepancy?

Alistair Law: No, they had not taken place. They were still proposals when I wrote to you, and they were still up for consultation.

Q77 Daniel Zeichner: There was a very detailed report by Sky News with lots of people speaking to them; you do not recognise that.

Alistair Law: A variety of different processes were ongoing. Some people had signed settlements and others were still under consultation at the time, but the broad proposals had not concluded their consultation process.

Daniel Zeichner: It feels a bit semantic to us.

Alistair Law: It is entirely in keeping with UK employment law approaches to ensure that you go through a proper consultative process. As I say, the changes were part of a set of wider changes that were made. It included UK colleagues. It also included colleagues in south-east



HOUSE OF COMMONS

Asia. All of it was designed to set us up for success when it comes to outcomes on the platform in relation to safety.

Daniel Zeichner: We will leave it there.

Q78 **Kit Malthouse:** I have to confess to being a bit confused, because you are all implying to us that your action on this sort of stuff is improving, yet the generally accepted trend, certainly for the three platforms to the left, is a move away from mediation. Alistair, your trust and safety teams have all gone. With X, there have been explicit decisions about abandoning policies on moderation. With Meta, there is an ideologically driven shift in policy to get rid of moderation and move towards us taking responsibility for what is on your platform in all its aspects. I want to ask a little bit about the evolution in decision making in all your organisations, because there has been quite a significant shift in the ownership structures and management of them all.

Can I start on X with Wifredo? You had a CEO in X who departed last year, so Mr Musk is now the de facto chief executive. He has amalgamated X into his wider empire, xAI and SpaceX. Who is making the decisions on this stuff? Where does it sit? Does it fundamentally go over his desk? We know that he takes a close hand in things. The whole controversy about turning Starlink off or on over Ukraine was down to his personal decision. Where does he sit in this decision-making tree?

Wifredo Fernández: Our safety team are the ones dedicated to enforcing our rules day in and day out.

Kit Malthouse: But they operate under instruction.

Wifredo Fernández: Sure.

Kit Malthouse: So where do the instructions come from?

Wifredo Fernández: We have written policies that they are enforcing against.

Q79 **Kit Malthouse:** Yes, but who is devising the policies? Is there a board that sits? Let us put it this way: there is no professional corporate figure sitting between these decisions and Mr Musk. If X wants to tweak its algorithm to amplify messaging that has a particular political flavour, would that go to him for decision?

Wifredo Fernández: No; you are maybe conflating his personal use of the platform and our operation of the platform. We have a dedicated safety team that develops and maintains our policies, and they are the ones tasked every day with enforcing against the policy.

Q80 **Kit Malthouse:** All right; forgive me. This operates on a human level. There is a human being somewhere in your organisation who is making a decision about how the software engineering should interpret data that is out there. Who is that person?



HOUSE OF COMMONS

Wifredo Fernández: We have engineering leaders and we have safety leaders, and they work in concert and co-ordination to enforce against our policy and develop new tools to enforce against our policy.

Q81 **Kit Malthouse:** Does it just emerge? Is there a committee? We are trying to get to some point of accountability for what is happening. You have been very professional coming out today, but you are telling us about something in respect of which, presumably, you have no decision-making role.

Wifredo Fernández: Right.

Kit Malthouse: So who is the person making the decision?

Wifredo Fernández: We have safety leadership who lead our safety organisation. We have a safety organisation and we have an engineering organisation within our larger corporate structure.

Q82 **Kit Malthouse:** They operate completely autonomously from the wider management of X.

Wifredo Fernández: No, they obviously work with our leadership on key decisions, of course.

Q83 **Kit Malthouse:** Who?

Wifredo Fernández: I think we all understand who the owner of the company is.

Q84 **Kit Malthouse:** Right. So he is directly involved in the production of these policies.

Wifredo Fernández: Yes, as any leader of a company would be.

Q85 **Kit Malthouse:** Right. Should we be concerned about one person deciding what is reflected off your platform into millions and millions of our homes?

Wifredo Fernández: For a particular company, decisions come down to one person, and that is the CEO or the leader of that company.

Kit Malthouse: Right, and those decisions will be affected by his view of the world.

Wifredo Fernández: I am not here to represent his particular views of the world, but I can tell you again—and I have said it before in this Committee—that our mission, our ethos and our operations are politically agnostic, and that is how we operate.

Q86 **Kit Malthouse:** Do you think we should all be concerned that just recently a jury found that he has misled investors?

Wifredo Fernández: I am not here to say what you should or should not be concerned about.



HOUSE OF COMMONS

Q87 **Kit Malthouse:** Right, okay. We have established that decision making goes right up to him on what the suite of these algorithms do. Thanks.

Rebecca, can we chat about the Meta situation? I understand there were proposals last year at your AGM for a board to be interposed to look at greater accountability and transparency, but that was vetoed by your chief executive's voting control of the organisation. What was the thinking behind that veto?

Rebecca Stimson: I cannot speak to that exact board decision, but to where you began with suggesting that we are moving away from moderation, I would challenge the premise of that question. I am happy to go into why. When it comes to how our policies are designed, we have established the oversight board, which you will be aware is an external, independent group of experts who look at our policies and challenge how we are enforcing them.

Kit Malthouse: I thought that since the departure of Nick Clegg that had all been abandoned.

Rebecca Stimson: No, it is still in place. We also have a policy forum within the company where across safety, product and a number of other teams, also with external experts in things like child safety and human rights, they scrutinise the policies that we are designing for the platform and the way that we are implementing them.

Q88 **Kit Malthouse:** Right, but in the end where does the decision lie?

Rebecca Stimson: The CEO is ultimately responsible for the company, but we have tried to introduce multiple layers, including the oversight board, of independent external oversight so that everyone can see how we are doing and where we have made mistakes. There are often cases where the oversight board disagrees with our approach, and we respect those decisions from the board.

Q89 **Kit Malthouse:** When a specific decision was taken to remove restrictions on political content, when you changed the content policies on immigration or gender that were eliminated from your policy platform, and when automated moderation systems were refocused to the very high end, that was all approved by this supervision process and signed off by your chief executive.

Rebecca Stimson: Yes, the decisions that you are referring to from January last year about how we altered the way we approached the most high-severity harms, and then other kinds of harms that fall below that, would have gone through that process.

Q90 **Kit Malthouse:** Right. Do you think the scrapping of the third-party fact-checking will have improved the quality of information on the platform?

Rebecca Stimson: We have not scrapped third-party fact-checking. It remains. We have one of the largest networks of third-party fact-checkers of any company still running globally.



Kit Malthouse: I thought you said you replaced it with community notes.

Rebecca Stimson: No, we are trialling community notes in the US, which is a different form of fact-checking. The aim still remains to get the best, most accurate information to people as quickly as possible. We think there are some advantages of the community notes model. It is borne out by some of the research, so we are trialling it there at the moment. Third-party fact-checking remains in place around the world while that process happens.

Q91 **Kit Malthouse:** I have a question about the two different platforms. On Instagram, you have, as you said, algorithms or whatever AI that stops nudity appearing, and it appears, certainly from my experience, to be quite effective, though people try to game it and there is an awful lot of semi-nudity. Why is that same technology not used on Facebook?

Rebecca Stimson: It is used on Facebook.

Kit Malthouse: And yet Facebook is responsible for the vast majority of child abuse reports that come through in the UK.

Rebecca Stimson: That does not always necessarily involve nudity. Your question was quite specific about the classifiers we have to find that. We have looked very closely at the reporting to NCMEC. We take that extremely seriously. Part of the reason we account for a large number of those reports is our effectiveness in finding it, passing it on to NCMEC and working with law enforcement where we come across that kind of content.

Q92 **Kit Malthouse:** Okay. Alistair, TikTok has had very significant ownership changes, certainly in the US, over the last few months. Who is taking the decisions on your algorithm now?

Alistair Law: I am here representing TikTok UK. We are legally established in the UK and obviously abide by UK laws. TikTok UK is owned by TikTok Ltd, and our CEO, Shou Zi Chew, who is based in Singapore, spends a lot of time over there. He is the ultimate decision taker. The US process that you are referring to has led to the creation of a joint venture in the US, so it is not a direct affiliate or subsidy of TikTok as a result.

Q93 **Kit Malthouse:** The decisions about the algorithm that are taken in the US are taken in a different forum from ones that are taken here.

Alistair Law: If you are going into the content recommender algorithm, that is licensed to the US joint venture and trained on US data over there. From a UK perspective, it continues to be trained on UK and European data. You asked questions about safety. We maintain community guidelines that govern what is and is not allowed on our platform. They have not gone through a huge amount of change. We are always iterating them, but they remain robust and aim to provide a safe and authentic experience.



Q94 **Kit Malthouse:** I am confused. What was the interaction between the UK recommendation engine, or content pushing agent, and the American one? Are they completely different? Does one feed off the other?

Alistair Law: It is licensed across to the American one, and the American one is then trained on US data as opposed to being deployed into data environments in the UK, Europe and elsewhere around the world.

Kit Malthouse: So that is just trained on data, and the actual system—the software engineering itself—is the same.

Alistair Law: Broadly, at the moment.

Q95 **Kit Malthouse:** In the same way that my TikTok account will show me different things from Adam's TikTok account because it is trained on different data, your man in Singapore is also individually making the decisions about how to tweak and turn up or down that recommendation engine.

Alistair Law: For the recommendation engine, there are a couple of things worth saying. The first is that the way that it is designed is to be very content agnostic. It presents you with a series of videos when you first use the platform that are usually our most popular videos.

Chair: I don't want to go through this again. We want to know who signs off.

Q96 **Kit Malthouse:** I am trying to get to the point of accountability. Lots of people are very concerned that for these profoundly important questions, which are the subject of dispute, alarm, concern and scepticism, the decisions lie in the hands, certainly for the three of you, of three rich, powerful men.

Alistair Law: From a TikTok UK perspective, decisions on both the algorithm and our community guidelines that set out what is and is not allowed are taken by our leadership team. From a trust and safety perspective, our global head of trust and safety is based in San Jose. He reports directly into Shou, who is our global CEO. When we make changes to the community guidelines, we do so on a consultative basis. Our trust and safety teams work with academics and partners around the world. We are transparent about the changes that we make—

Kit Malthouse: Yes, but in the end it has to go for approval over his desk, right?

Alistair Law: Yes, correct.

Q97 **Kit Malthouse:** Okay, right. The fundamental problem we have is that the prejudice, judgment—call it whatever you like—of these three people is basically appearing in every home, every day, all the time. You can talk to us all you like about the systems and what have you, but in the end you tweak it to their desires, and that must be of significant concern. Is



HOUSE OF COMMONS

there no move in any of your companies to diffuse that individual influence and find some way to make it more democratic, for want of a better word?

Chair: Corporate governance.

Alistair Law: Under the OSA, there are governance requirements, so decisions on safety, decisions on our community guidelines and how we then enforce against them are indeed taken in a collective, collaborative way by a variety of different experts across the company. Ultimately, the driver for our output here is what users want and what experience we are trying to create, for instance.

Kit Malthouse: Just to finish, because I am conscious I am out of time—

Chair: We are running out of time.

Kit Malthouse: You have talked to us about the improvements you have made in your systems. Is there any chance that you could write to us about improvements that you have made in your internal governance to limit or control the influence of these three very powerful individuals? I am sorry, Zoe, that I have not got on to you, but you are in a slightly different position. The other three, as far as I can see, have moved away from content moderation, whereas you are broadly static. It would be helpful to understand what their intention is.

Chair: Thank you very much, Kit. The British public have a right to know how such influential companies are taking decisions that affect the content that they and their children see. We have to move on now, as we are coming on to the last 28 minutes, and I want Lauren to take us forward.

Q98 **Dr Sullivan:** Thank you, Chi. The Government are consulting on options for protecting children from online harms, and that consultation opened on 2 March and closes on 26 May. This is a simple yes/no answer from each of you. Should the age of children who are allowed on a social media account be raised? Over to you, Rebecca.

Rebecca Stimson: I know you want a yes or no.

Dr Sullivan: I do. I am coming back for some more.

Rebecca Stimson: Moving the age does not solve the problems that the Government are trying to solve. We have teen accounts where we have defaulted all our under-18 users into a much more restricted experience. We are completely with the intention of ensuring that young people have age-appropriate experiences and giving their parents a high degree of control. To your question on simply moving the age, we need to have a broader conversation, which is what the Government are trying to do in the consultation, about how best to make that effective and implement it at scale.



Dr Sullivan: That is kind of a no.

Rebecca Stimson: It's kind of a no.

Wifredo Fernández: We have no business in kids. We are not trying to build a business around kids. Globally, less than 1% of our user base is under the age of 18 and over 13. In the UK, that is less than half of a per cent. We do not have a major dog in this fight, if you will. There is a legitimate question—and this is a policy question for every jurisdiction—of how you gate access to information on the internet. With X being a unique platform for understanding what is happening in the world, it is a relevant question. We are sort of neutral at the moment.

Alistair Law: I would say similar. It is for policymakers to decide in this area. I completely understand the level of concern that exists, and that exists across concerns to do with harmful content, concerns to do with level of screen time, concerns to do with wellbeing. We agree that if you are 15, you should have a very different experience online from if you are 25, and that is why we have 50 preset safety features for our under-16 accounts. The OSA regulates about 150,000 services, from Ofcom's figures. If the policymaker view is that that long tail of 150,000 services where people may be spending their time needs significant action to be either brought into an age-appropriate experience or prevented altogether, that is a decision for policymakers.

Dr Sullivan: Okay. So that is probably a "No," and back to us.

Alistair Law: I think I would say it is a neutral.

Dr Sullivan: Neutral, okay. Zoe, what do you think?

Zoe Darmé: I do not work on any social media products, but I can give you my own opinion.

Dr Sullivan: YouTube?

Zoe Darmé: YouTube is a video library—a video-sharing platform. It lacks many of the social features that traditional social media companies have produced. For me to answer this question, because I work primarily in—

Dr Sullivan: It is a yes/no answer.

Zoe Darmé: For Maps, which is regulated as social media under the DSA, I would say no, because I believe that children should be able to use Google Maps.

Q99 **Dr Sullivan:** Okay. Would you describe your platform as safe for children and young people, Alistair?

Alistair Law: Yes. As I say, we have under-16 age-appropriate experiences. For example, if you are under the age of 16, you do not have access to direct messages. Zoe talked about certain elements of



HOUSE OF COMMONS

social media. The under-16 experience on TikTok looks quite a lot like YouTube, because it is video-based comments. You are set to private by default. We have, as I say, about 50 different preset settings for people under the age of 16.

Q100 Dr Sullivan: Okay. What is the minimum age that a young person can join TikTok?

Alistair Law: Thirteen.

Q101 Dr Sullivan: If we know that over 80% of 10 to 12-year-olds have an account with at least one platform, some of those would be yours. What are you doing to ensure that that 13 cut-off is enforced?

Alistair Law: We have a multi-layer approach. First, if your age is set in the app store to be 12 or under, TikTok will not even appear for you to be able to install it. When you try to install it, you are presented with a neutral age gate set to today's date and asked to put in your birth date. If you put a date in under 13, it will block you from reapplying. Obviously, we know that people will try to circumvent that, so we put people into an under-18 content experience, and then we will use technological estimation to assess how old they are on the basis of their behaviour, along with user reporting and dedicated teams aimed at finding people under the age of 13 through signals like what they put in their bio or comments that they make. We report every quarter on the number of under-13 accounts that we enforce against and take people down, and we will continue to do that in a transparent way.

Q102 Dr Sullivan: Okay. Over to Rebecca. Would you describe your platform as safe for children and young people?

Rebecca Stimson: We have teen accounts where, as I have said, we have defaulted users under the age of 18 into a restricted experience.

Dr Sullivan: Is it safe?

Rebecca Stimson: Ninety-seven per cent of parents and children have chosen to remain in that experience, so we are taking that as an indication that they are finding it extremely helpful. We are very conscious that there is still harm. This work is never done. I would never go as far as to say that there is a zero risk of anything harmful. There are real challenges around age assurance and ensuring those age-appropriate experiences.

Q103 Dr Sullivan: Why do you think that there is this growing support for an outright ban, to protect young people and children from your platforms, if what you are telling me is that it is safe and we do not need to make any changes?

Rebecca Stimson: Evidence of the support is really mixed. I am sure you are aware of that. I watched your previous evidence sessions about the ban. There was evidence to show that parents were not in favour of it necessarily. Half of them in various studies have said they turn it off.



HOUSE OF COMMONS

What we hear consistently from parents globally about their concerns is three things. It is about contact with their children, content that they are seeing and screen time. We focused on addressing those through teen accounts, so that parents have greater levels of control and insight and teens are defaulted into a much safer experience.

We have not seen universal support particularly for a ban. I completely understand why it is a big topic of discussion. People are clearly very concerned about the impact of social media and other types of apps on their children. You mentioned the stat about the number of under-13-year-olds; I think that came from the recent Ofcom work which, as you say, showed that 70% of young people are on YouTube. There are lots of interesting questions about the ban and the scope of that ban.

Alistair Law: As I said earlier, the variety of different concerns that people have absolutely resonate with everyone as parents, whether it is concern about harmful content, screen time or wellbeing. We invest heavily in trying to create an age-appropriate experience and address each element of that by giving people agency. We have a 9 pm cut-off for notifications. We have a 10 pm total screen takeover for under-18s, where a message will come up and say that maybe it is time to move away from the app and take you through a meditative breathing exercise. I appreciate that the experience people have online goes over the 150,000 sites that Ofcom is regulating. From a parental perspective, I completely understand that there is a high level of confusion.

Q104 **Dr Sullivan:** You mentioned at the beginning that it is based on the phone set-up whether a young person can even access TikTok.

Alistair Law: That is one element of a multi-layered approach.

Dr Sullivan: Should parents play a bigger role in making sure that their phones are set up for their young people?

Alistair Law: We think that our responsibility is to make sure that our platform is as safe as it possibly can be and as effective at being able to identify people's age, even absent those steps. We try to raise awareness with parents of the tools that we have available.

Family pairing is our parental control set of tools. If you are over the age of 35 and join the platform, you are automatically, whether you are a parent or not, shown some information around family pairing and how to set it up. At one point, we even gave a notification to every over-35 user across the country of the features and functionalities that we have, because we recognise that parents are looking for support and the ability to tailor the experience to what they and their teen may think is particularly appropriate for them. It has features such as being able to set blocks of times away from usage and so on. We want to support parents in a partnership, but we completely accept and drive investment into our responsibility to make the platform safe and age-appropriate.



Q105 Dr Sullivan: Okay. I want to return to a piece of research that was carried out by the National Education Union. I am jumping ahead a little bit, but I thought it is relevant for this point, in particular for Instagram and Meta. The NEU conducted an experiment where it set up five different accounts as 13-year-olds and, after going on Instagram for half an hour every day for a week, recorded clips of the feed that it saw. We were unable to share this today because of the amount of violent, misogynistic, self-harm and extremist content. I have seen it. It is appalling. We cannot show it today, but that is being fed to 13-year-olds, under-13-year-olds—whoever. Can you say again, Rebecca, how you are making it safe for our young people to see this content?

Rebecca Stimson: I am aware of that report. Clearly, we are going to look at that very closely and see what happened in that case. Sometimes these things can be difficult if people set up accounts to test the systems that do not behave like normal users would. I am not trying to make an excuse, but the robustness of some of those findings is interesting to look at. We are absolutely looking at what was found.

As I mentioned, we have rolled out teen account protections that give default younger users a very different experience and a very much safer experience. Ninety-seven per cent of parents and children have remained within that. The feedback that we have had is extremely positive globally from the parents who have seen what we have built there, including all the tools that help them manage the experience. We are working very hard.

We are one of the founders of the OpenAge Initiative. Recognising that age assurance is an imperfect science, as Ali said, we collaborate with companies like Yoti. We have AI detection. We ask for verification. The OpenAge Initiative is looking at interoperable keys on someone's device that would help age assurance across all platforms. It could move around.

I recognise the content that you are referring to. We will look at it very closely and take that very seriously.

Q106 Dr Sullivan: Can you write with the actions?

Rebecca Stimson: We will, absolutely.

Q107 Samantha Niblett: I want to talk about money. How much revenue do you generate from child users in the UK?

Rebecca Stimson: I do not have a figure, but it is extremely low because our business is an ads-based business, and our ability to advertise to children is extremely restricted by both the law and our own policies. They are not a revenue generator for us.

Q108 Samantha Niblett: Before I come back and ask the same question to the rest of the panel, Meta deactivated 554,000 accounts following the social media age restrictions in Australia. With you saying you do not



HOUSE OF COMMONS

make much in the way of advertising revenue, how has this impacted the revenue in Australia? That is less of a “finger in the air” question.

Rebecca Stimson: As you said, we removed 500,000 accounts in response. We are working with the Australian authorities on an updated set of data now that the ban has been in place for a while, and we can see how people are behaving and what they are doing. I do not have a figure to share, but I would be happy to follow up with some more information.

Q109 **Samantha Niblett:** I appreciate that. Wifredo, I know you said you do not have many young people using your platform. I am quite pleased. As a mum of a teenager, I asked her group of friends, “Do you guys use X?”, and they said, “No, just for porn.” Do you make any money off the back of young people that you can identify in the UK?

Wifredo Fernández: You cannot specifically target 13 to 17-year-olds, though they may be included as part of a broader audience that an advertiser may include—for example, 13 to 30 or 13 to 50. There may be some revenue based on general advertisements.

Q110 **Samantha Niblett:** How would you work out how much revenue that is? Do you know?

Wifredo Fernández: I am happy to investigate that. It is a matter of account impressions and which ads they sell.

Q111 **Samantha Niblett:** Okay; thank you. Ali?

Alistair Law: Mine is a very similar answer to Rebecca’s. We do not target personalised advertising to anybody under the age of 16, and therefore it is minimal, negligible levels.

Q112 **Samantha Niblett:** Just out of interest—sorry, Zoe, it feels probably like a less relevant question to you, which is why I have not asked you—what is in it for you to suck kids in, then? Is it basically that it is making yourself ubiquitous in their lives so that when they are older they are addicted to your platform so that you can make money? You are big, corporate organisations that are not here to end world starvation or aid world peace, because we are seeing the absolute opposite happening. You are telling us that having children on your platform is not a bad thing, but you are also saying you are not making money, so I am just trying to understand what is in it for you.

Alistair Law: From my perspective, it reflects the answer that I gave to Dr Sullivan earlier. If policymakers make a decision around 16, it is definitely not something that we are going to oppose. One of the things that we are trying to highlight—

Chair: That is not Samantha’s question. Her question is: what do you gain from having children on your platforms?



HOUSE OF COMMONS

Alistair Law: Children on the platform gain access. I said that 0.7% of content was violative. That means that 99.3% of it is not violative. It is valuable. It is people finding community and finding connection.

Q113 **Chair:** Samantha's question is: what does your company gain? Sorry Samantha, but we are not getting an answer here.

Alistair Law: We bolster our community overall; people have experiences that are valuable to them and they can find information. At the same time, the UK Government are talking about dropping the voting age to 16. Platforms like TikTok are ways to be able to view the world and engage with the world.

Q114 **Samantha Niblett:** Would any of you say that your engagement with young people is altruistic? You are doing it for the betterment of society.

Rebecca Stimson: Clearly, it is in our interest to attract users to our service, I guess, but in order to do that the aim is to provide them with the most valuable and enjoyable service that we can, and that is for the parents and young people who, as Ali has just said, use our platforms in a whole array of incredibly positive ways. It is obviously our responsibility to also ensure that we are addressing harms that can happen online.

Q115 **Samantha Niblett:** I will rattle through because I am conscious of the time. I am really keen to ask a question to do with revenue generating. Lots of organisations make money by having a presence on your platform. As MPs, whether we like it or not, we have to use Facebook, even when we feel a bit grumpy with you.

I will use an example. In my constituency, Cody Chetwynd runs a 1950s American diner and has over 32,000 followers. People do not use the website; they go to Facebook pages. Her page there was taken down, and all she was told was, "You violated community standards." She couldn't get in contact with anybody. My colleague and friend, Adam, is a big fan of the restaurant as well. Out of pure desperation, they were pleading for help on Instagram, so I stepped in. It is only because I know a couple of people in Meta that I am able to plead for help, but I am still being told, "This might take a while."

Can you even imagine what it would be like on the flip side for us to say to social media platforms, "We are just going to switch your platform off because you violated community standards," with no explanation? When we are talking about how, as a country, we want to aid economic growth, what can you do to make it better for businesses that are using your platform, and be more transparent with them, so that their businesses are not put at risk because they have suddenly lost an audience of 32,000 followers?

Rebecca Stimson: I do not know about that exact example. If it is still unresolved, I would be very happy to look into it specifically for you. It is a real challenge. I know how people can get very frustrated, because obviously we totally recognise the direct impact that would have on a small business. We try to invest in the right systems to respond as



HOUSE OF COMMONS

quickly as possible and give people some additional context. It does not always work perfectly, and I accept that. Please do carry on flagging up to me and my team if there are specific examples, and we are more than happy to have a look at them.

Q116 **Chair:** I just think that is absolutely unacceptable. You are saying that politicians should be flagging to you when small businesses, which are the lifeblood of our country economically and the driver of growth, have their entire revenue stream taken offline. You are accepting that your systems do not work.

Rebecca Stimson: Our systems are not completely perfect. They deal with the vast majority of these kinds of account problems very swiftly. If there is something where it does not work perfectly, all I am saying is that we are happy to look at that and make ourselves available to the Committee to solve those issues.

Chair: Okay. Thank you for that.

Q117 **Martin Wrigley:** I want to talk to the three of you, because Google Search is somewhat different, although I would differ with your view that YouTube is not social media. I think we will find that it is. In terms of the three of you, you run businesses predicated on attention due to inflammation and excitement, inciting that connection or addiction. You came in this morning really complacent about all the issues that we were talking about, and you started off by saying, "Everything's fine." We have gone through and demonstrated a number of different occasions when things are not fine on your platforms, no matter how much cognitive dissonance you put into it.

Your algorithms, recommended content and publication of content is curation, which in my view turns you into publishers. You have responsibility, and are taking responsibility, for the content that is going on your platform. What you are doing is not working, because the problem is getting worse. You are continuing to fan the embers of misinformation and disinformation. Sometimes you will send the fire brigade when the fire gets too big. Going forward, what are you going to do differently? Rebecca?

Rebecca Stimson: Is there a specific issue that you are referring to?

Martin Wrigley: I am talking in general. I do not want to see you sending a fire engine out every time. I want to see you stopping this systematically. What are you going to do differently? What you are doing now is not working.

Rebecca Stimson: First, I am sure that none of us is intending to come across as complacent. We are trying to explain what we are doing and some of the ways in which these things are imperfect. Certainly, the evidence that we publish every quarter in our transparency reports about how we are doing against harmful—



Q118 **Martin Wrigley:** What are you going to do differently? We have heard what you have been doing. What are you going to do differently? What is next? How are you going to fix this? Because what you are doing is not working.

Rebecca Stimson: We continue to make very substantial investments in safety and integrity across our platform. We continue to have thousands of people in the company dedicated to focusing on this issue. One big frontier that we are looking at is the advent of AI and some of the challenges that that is posing. As I have mentioned already, we are working collaboratively with other industries to try to identify some of the technical challenges that are happening there.

Q119 **Martin Wrigley:** Time is short. You are saying more of the same; that is not working. X, what are you going to do differently?

Wifredo Fernández: Simply put, give users more choice and control over what content they see and how they see it. That is why we are experimenting with content filters on the “For You” timeline. We have two timelines, one chronological and one that is a recommendation algorithm. We give people topical choice to be able to toggle on or off between topics. We will eventually move towards a system that gives users a lot more control and be able to facilitate it.

Q120 **Martin Wrigley:** You are saying it is not your problem; it is the user’s problem. TikTok, what are you going to do differently?

Alistair Law: A combination of both of those things: a greater level of managed topics, similar to Wifredo’s answer, and then really double down on automated and AI moderation. A couple of years ago, AI was very good at nudity detection, but it could not tell necessarily whether a knife was being used in a cooking video or in an attack. As we have increased the level of sophistication, it is able to do that, at speed and with high levels of accuracy. That is the way in which we will drive continued improvement at scale.

Q121 **Martin Wrigley:** To give you credit, you have talked about more things that you are doing than the other two. I have one final question. All your platforms suffer from huge numbers of non-people users—bots—that post all sorts of things all over the world, which are making your lives harder. What are you doing about that, please, Rebecca?

Rebecca Stimson: We remove up in the 90 percentiles of fake accounts that appear on our platforms, and that can include bots. We are quite effective at finding and identifying those and removing them.

Q122 **Martin Wrigley:** It is clearly not working, so, again, you need to do something differently. What about at X?

Wifredo Fernández: Spam and platform manipulation is top priority, so we are going to continue to intervene there. It is an adversarial space. As soon as interventions are implemented, soon thereafter there are new and novel ways. It is a constant challenge, there is no doubt about it.



Q123 **Martin Wrigley:** So no new ideas. Carry on doing what you are doing. TikTok?

Alistair Law: We have policies directly against fake, inauthentic accounts that start from the idea of spam all the way through to co-ordinated influence operations. We have a variety of both off and on-platform work that we do in relation to that. We would love to be able to get additional levels of signals for that. One thing that I spoke about at the Foreign Affairs Committee recently was whether there are broader levels of value chain. Some of the things we do on a cross-industry and cross-Government basis for things like terrorism and illegal content. There are lessons to be learned for other types of harm, too.

Martin Wrigley: I am not seeing anything different there, either. Thank you, Chair.

Chair: Thank you very much, Martin. We are running out of time. I am going to go to Kit briefly and then Emily.

Q124 **Kit Malthouse:** Thank you, Chair. I have a very quick follow-up to Rebecca. A BBC report recently said that internal Meta documents said that the “set of financial incentives our algorithms create does not appear to be aligned with our missions”. What does that mean? Is it still the case? Where in the ranking of importance does revenue maximisation sit in your algorithm?

Rebecca Stimson: Those are from internal documents that are now several years out of date. They do not reflect the algorithmic changes, which I referred to in an earlier answer, that we made back in 2018, and also some last year about our ads algorithm. Our users and our advertisers are absolutely clear that they do not want to see harmful content online, so it is in our business interest to make sure that the algorithm is finding and removing that as much as possible in order to make people have the best experience they can on our services.

Q125 **Emily Darlington:** I have a very quick question. YouTube Kids is a platform marketed towards children. Could you tell me what the BBFC rating would be for that site? Is it U, PG, 12?

Zoe Darmé: I am sorry that I cannot answer your question. I do not know the—

Emily Darlington: I have been asking for a year. Could you please write to the Committee about what the BBFC regulation is for the content that appears on YouTube Kids and that is being marketed to our children?

Zoe Darmé: We can certainly respond to your question. I apologise that you have asked my colleagues, including YouTube colleagues, in the past and have not gotten an answer.

Emily Darlington: Thank you.

Chair: Thank you very much. I hope you will be taking away from this



HOUSE OF COMMONS

session the sense that the Committee feels quite strongly that there is a level of complacency in your response to our questions and in the evidence you are giving. We have had concerns expressed about the governance and the amount of control that individuals have over the algorithms and how your platforms work. We have had a lack of future-proofing. We would appreciate you writing to us about that. The basic fact is that all the work that you tell us you are doing on online harms and to make your platforms safe in this country is not working. That is the consensus of the Committee. That is the consensus of most of the British people.

Emily Darlington: For women, and minorities.

Chair: Yes, but for all kinds of different groups. What we would like you to come back to us with—and we expect to see you in a year's time, if not before, to see progress and get an update—is how your products are going to be safe for the British people. The alternative, if you cannot make it safe, is that we need further legislation to make it safe. The first duty of any Government is to protect their citizens.

Thank you very much. I know it has been a lively discussion, as we say. I look forward to seeing you again. We will look forward to hearing your responses to the particular questions that you are going to come back to us on.