

Science, Innovation and Technology Committee

Oral evidence: Social media, misinformation and harmful algorithms, HC 441

Tuesday 18 March 2025

Ordered by the House of Commons to be published on 18 March 2025.

[Watch the meeting](#)

Members present: Chi Onwurah (Chair); Emily Darlington; Jon Pearce; Steve Race; Dr Lauren Sullivan; Adam Thompson; Martin Wrigley.

Culture, Media and Sport Committee member present: Paul Waugh.

Questions 180 - 248

Witnesses

I: Dr Karen Middleton, Senior Lecturer, Marketing, University of Portsmouth, and Adviser to the Conscious Advertising Network; and Phil Smith, Director General, Incorporated Society of British Advertisers.

II: Dr Eirliani Abdul Rahman, Online Safety Advocate and former Trust and Safety Council Member, Twitter; and Lyric Jain, CEO, Logically.

Written evidence from witnesses:

- [Dr Karen Middleton](#)
- [Dr Eirliani Abdul Rahman](#)
- [Lyric Jain](#)

Examination of witnesses

Witnesses: Dr Middleton and Phil Smith.

Q180 Chair: Good morning and welcome to the third session of the Committee for Science, Innovation and Technology's inquiry into social media, algorithms and the spread of misinformation. I welcome Paul Waugh from the Committee for Culture, Media and Sport, who is guesting on this entire inquiry, and I welcome our two witnesses for the first panel.

I will start by asking you both a question and giving you the opportunity to introduce yourselves as you answer that question. The Committee understands that up to 98% of the revenues of social media platforms Meta, TikTok, X and Google as a search engine comes from advertising. Advertising, therefore, is a critical part of the business model, the biggest and most important part of the business model of social media companies. My first question to you is: do the business models of social media companies rely on or encourage the spread of harmful content?

Dr Middleton: Thank you, Chair. Thank you very much for inviting me to this inquiry and for giving me the opportunity to talk about my research in an area that I feel very strongly about and have been looking at for the last couple of years. I am Dr Karen Middleton, senior lecturer in marketing at the University of Portsmouth. I am also a voluntary adviser to the Conscious Advertising Network and have been with them for two years.

To answer your question about the underlying business model of social media and how that is fundamentally advertising, I would give an emphatic yes, but I would also say that it is not just social media; it is effectively the internet in broad terms. If websites use advertising, that can be something that enables them to exist. There is a lot of money to be made from online advertising—estimates at \$1 trillion for 2025. Yes, in answer to the question, I agree with that.

You ask how that business model spreads misinformation and online harm. There are four areas to look at. The first is the algorithms that social media and other websites use to serve content and to serve advertising to the users. Inherently, content and people's interest in staying online is very often driven by sensationalist content, content that is engaging or content that is emotional. That very fact, a fact of human nature, drives the engagement. The algorithms have no differential between deciding whether something might be harmful and if it has spread, and deciding what the nature of that content is and the effect it might have. In themselves, those algorithms prioritise harmful content by definition.

Moving on from that in itself, which we could call the attention economy, which has been in existence since the advent of newspapers for many years, the second point around the fundamental business model of advertising online is that what has emerged from that ecosystem is the



HOUSE OF COMMONS

proliferation of advertising fraud. There are a lot of websites used for clickbait to make money from advertising. Those websites are often fraudulent, often based on fake news websites, as we saw with the Southport riots and Channel3Now. Within that context, ad fraud has the same lack of transparency as the algorithms and is part of that harmful system. Those fraudulent advertising networks may often fund criminal activity, but they could be seen as the disinformation economy. It serves the purposes of fraudulent advertising websites because of that tendency to proliferate contentious, enraging content to get clicks and to benefit from advertising.

A third point is something that has already been mentioned in previous meetings of this inquiry, in particular by Marianna Spring. These algorithms, in their scale and their prioritisation of this kind of content, as well as spreading harmful content and misinformation, very much normalise this kind of content. The media in itself cultivates certain societal norms. There is a type of credibility associated with things that are in writing, that are online. In essence, it is publishing of content that is often seen by users as true, because it is so commonplace and it is spread online.

Underlying that, my fourth point is about the major platforms. There are obviously some dominating platforms; Google and Meta dominate the market, but they face little scrutiny. They are not regulated in the same way as other media were previously, before the advent of online media. Those things set the context, I think.

Q181 Chair: Thank you very much, Dr Karen Middleton. Phil Smith, that is an unequivocal confirmation from our first witness that the business models social media companies rely on encourage the spread of harmful content, and advertising is a fundamental part of that. What is your response?

Phil Smith: Thank you, Chair. I am sorry that I am not able to be with you in person. My name is Phil Smith. I am the director general of ISBA. We represent the UK's leading brand advertisers.

With qualification, I would give you some agreement that, in general, where maximum audience and attention are rewarded, there are attendant risks, as there are with other media. It is especially so for social media, where content is generated at vast scale and it is served on the basis of engagement. Most highly engaging content is harmless—think of cat videos—but some is clearly not. Users with harmful interest can connect readily with victims or with collaborators. That risk becomes particularly evident in fast-moving crises like last summer's riots or after the George Floyd murder five years ago in the United States. The platforms' historical lack of liability for what they carry has compounded the issue, I believe.

We should acknowledge though that, over the last seven years, there has been huge investment by the major platforms in offsetting the risks, employing tens of thousands of people in content moderation and



HOUSE OF COMMONS

building automated detection, as well as in transparency reporting and audit. This has been encouraged by brand advertisers. We should note that there are some smaller platforms with fewer issues in this regard. Pinterest, for example, claims to prioritise positivity in its content recommendation and there have been fewer concerns about harmful content reported from that platform. Scale, of course, may be a factor.

Q182 Chair: Phil Smith, as a representative of the advertising sector, you agree that attention maximisation will drive harmful content. Are you saying that the measures some of the platforms are taking that you described help to address that? What more do you propose needs to be done?

Phil Smith: There are greater attendant risks with social media of harmful content spreading. I believe the major platforms have taken very large steps and made investments to offset those risks, but the question of whether enough is enough is a difficult one to answer. With some of the most severe harmful content, zero tolerance is all we could allow.

It is encouraging that the Online Safety Act, which we have supported since its inception, is now coming into its first major tranche of enactment. We believe the Online Safety Act needs to be seen as a positive force in forcing platforms to remove illegal content, as it intends to do, very promptly.

Chair: We will come on to discussing the Online Safety Act and its impact. You talk about illegal content. This Committee is most concerned with harmful content and misinformation. Martin, do you have a point to raise?

Q183 Martin Wrigley: Thank you, Chair. You have both touched on this, talking about the historical lack of accountability for the systems we refer to as platforms. There was a big debate in the early days of those systems as to whether they were platforms, thus not accountable, or publishers, thus accountable. Has the way they work today, with recommendation engines, with curation and effective editorial selection, actually moved them from the realms of platforms to publishers?

Dr Middleton: I would say yes. I am familiar with that question. I must admit I am not really clear on why they would not be considered to be publishers, considering this information is pushed out at scale and is in the public domain. Perhaps Phil would say something about that. I am confused as to why they wouldn't be considered to be platforms rather than publishers.

Q184 Martin Wrigley: It was early days regulation decisions that took them to become platforms and thus not liable for the content they had on their neutral platform, as opposed to a publisher who has editorial choice and, thus, liability for what they publish.

Phil Smith: In particular, section 230 in the United States gave them zero liability for what they carried. I am not sure whether the distinction



between publisher and platform is a useful one here. What is really useful is whether or not they are liable for what they carry, and that is clearly something the Online Safety Act is intending to address. That is why we have been keen for oversight and regulation, going back to 2018.

Martin Wrigley: Thank you.

Q185 **Chair:** You mentioned the Online Safety Act, which deals with illegal content. It is not as clear, in its implementation or in the details of the Act, what its approach is to misinformation, which is what we are considering here. There is lack of clarity about the algorithms by which the social media platforms distribute content and the way in which the digital advertising processes, which drive the financial incentives for the platforms, are designed. Is it fair to say, Karen Middleton, that the technology sector has designed digital advertising processes to be complex and opaque, with minimal human oversight?

Dr Middleton: I am not sure there is an intentional drive to do that. It is probably more to do with how it has emerged, how the technology has emerged and the need to do things at scale. There is a very complex and opaque supply chain nevertheless. The harms we see from this process could be described more as unintentional outcomes, but, nevertheless, unintentional outcomes that drive misinformation and harmful content are, in and of themselves, profitable. There is no drive to change them for those who benefit from it.

In the digital advertising system, there are many layers. It is broadly in four layers, where advertisers liaise with demand middlemen, then supplier middlemen and then the websites or platforms where the advertising is placed. That in itself obviously creates opacity. It also means many actors in the supply chain do not have awareness of where their advertising is placed; they do not have full control of the system because of, I guess, the scale and the different options and the fact that programmatic advertising itself relies on real-time bidding of engagement. There isn't opacity for those reasons.¹ This is probably something we will go on to talk about: in the advertising ecosystem, especially with the smaller companies and those in that supply chain, there is a huge level of uncertainty and sense of responsibility over that problem. However, there remains a lack of control because of the disjointed nature of the supply chain and the lack of opacity and the requirement for that to be made clear by the different levels in the system.²

Q186 **Chair:** Should the Government act to require greater clarity? What would you recommend or propose?

Dr Middleton: One of the key levers for that would be to recognise the problem and—I am not sure if I would use such a strong word as

¹ Clarification from witness – "There isn't the transparency there for those reasons."

² Clarification from witness – "...the lack of transparency and requirement for..."



HOUSE OF COMMONS

“malpractice”—the financial incentive. As such, I suggest it should be considered by the Financial Conduct Authority as well as other regulators or organisations, and the Government, within that realm.

We talked in previous briefings about the Advertising Standards Authority, and I don’t know if we will go on to talk about that in more detail. The Advertising Standards Authority is a self-regulatory body; they are not Government-funded. They do lots of positive work obviously for the industry in helping to regulate, but it is much more about the content of the advertising itself, rather than anything about the ecosystem and how that might drive harm, which has generally been recognised as a problem only in recent years. Perhaps there is a situation where it has not caught up.

Q187 Chair: Thank you very much. That is very helpful. Paul Smith, we have heard that how adverts are placed where is somewhat opaque. Do you agree that advertisers are often outraged to find out where their ads end up online? How should that be addressed?

Phil Smith: If I may, Chair, I would like to go back to the earlier question to give a bit more clarification. When you talk about complexity and opaqueness, I would in general agree, but you have to be specific about which parts of the sector you are talking about. When you talk about the walled garden platform environments, like Facebook, X or Instagram, for example, advertisers can buy directly from those platforms. Of course their processes are automated and there are questions about where ads appear, but there are services that allow you to see where your ads have appeared.

The most opaque segment in the market, which the previous speaker talked to, is programmatic advertising on the open web on websites. That is particularly complicated and opaque because, as has been said, there are many steps and actors in the supply chain. You have agencies, you have demand-side platforms and supply-side platforms before the money gets to the publisher. ISBA has conducted studies with advertisers and publishers over the years that demonstrate the complexity and the flows of money. In our last study, in 2021, as an example, the average number of websites that were included in a campaign was 9,000; a lot of advertising was going to a lot of websites with a lot of lack of clarity about where it was going. Going back to the earlier point, the made-for-advertising websites proliferate where they have poor or false content designed to attract audiences cheaply.

To your point about brands and advertisers and their reaction, of course they are hugely concerned. ISBA has led multiple representations to platforms over the years—for example, when it was found in 2017 that brands were advertising on terrorist propaganda videos and sharing revenue with their creators. We have participated in the Global Alliance for Responsible Media, which was the voluntary effort seeking greater levels of transparency on user safety and control over ad placement. As you are aware, this activity has now been discontinued as a result of a



lawsuit from X claiming that the alliance was collusive in nature, which is being contested by the World Federation of Advertisers.

We have always been clear that advertisers have to make their individual spending decisions based on their own priorities, and that advertisers should not seek to influence content, but should demand clarity on policy, and assurance on the effectiveness of the policy implementation. That is what we have sought to do in the individual clarification and guidance we give our members on a regular basis.

- Q188 Jon Pearce:** Phil, I want to go slightly further into that. I have been looking at a study done by the Centre for Countering Digital Hate, which looked at harmful eating disorder videos being recommended to children as young as 13. Those videos included, "What I eat in a day," a video showing extreme calorie restrictions; "Thinspiration," a video presenting skeletal bodies as inspiring; and "ABC, the anorexia bootcamp," a diet video of zero to 500 calories and extreme diet. What the Centre for Countering Digital Hate found was that brands, very senior brands like L'Oréal or Ralph Lauren, were being shown next to some of that harmful content. Should we be doing more to protect children online, and how should we be doing that? Is there a responsibility for the advertisers? Is there more that we should be demanding from the platforms on how they protect young children online?

Phil Smith: The first thing is that we have always said more needs to be done to ensure that harmful content does not appear on platforms or on the web at large. The regulation of that is a matter for Government. Advertisers definitely want to see a break in the link between their advertising and the funding of that content and have fought hard to make sure that they have more control over where ads get placed.

Of course, content like that is unacceptable. The appearance of advertising next to it will not be acceptable to the advertisers either. What I would suggest is that it is a collective effort, if it is to get to a place where we see no recurrence of it.

Jon Pearce: Thank you, Phil.

- Q189 Emily Darlington:** Phil, I am going to push you a bit. You say that it is up to Government, and we fully accept it is up to Government to regulate in this area. What is it that Government could actually do to give transparency to advertisers so that their content does not appear next to harmful content—not just Meta, but TikTok and Instagram as well—given the algorithm pushing this eating disorder material? Those are proliferate sites.

Phil Smith: The platforms themselves would say that they provide a lot of control to advertisers over where their ads appear these days. In particular, Meta is introducing, and continues to introduce, controls over placement based on brand suitability. Those controls are clearly not perfect, and we see issues coming through where both the content is



HOUSE OF COMMONS

unacceptable and where the placement from the point of view of the advertiser is unacceptable. There are services available, which advertisers often use, that examine in detail where their ads have appeared next to. Those are third-party organisations, some of which have been independently verified.

Q190 **Emily Darlington:** Sorry, I don't think that answers my question. Apologies. You said that the Government should do more to protect harmful content and to bring transparency. I am trying to get to the bottom of what you think Government should do.

Phil Smith: First of all, we think the Online Safety Act needs to be allowed to take effect. A major tranche of its powers come into effect this week. That will work on illegal content. When it comes to where advertisers place their ads, we are currently not asking Government to do anything.

Q191 **Chair:** That is a really important point Emily is making, Phil Smith. If the sector, the industry, is not asking for any change when it comes to transparency or control of the placements of ads, the advertisers have to accept their responsibility when those adverts are funding harmful content, which we see is happening.

Phil Smith: There are some distinctions to be made. In some settings, advertising leads to a revenue share with the creator. It is clearly unacceptable for any advertiser for that to be happening and there has to be responsibility for the platforms and the advertisers themselves. In other circumstances, you see advertising appearing in a newsfeed environment where harmful content also appears. Advertisers do not wish to see that happen either. If there is funding of the platform, it is not going to the creator necessarily. That link is something we need to be careful of when trying to make those connections.

Of course, advertisers want to see their ads appearing in the right places. They work hard with the platforms to drive better standards of clarity around that. They work hard with independent third parties to check where those campaigns are going. When it comes to the advertiser community, one thing we should remember is that while brand advertisers will be spending large amounts of effort and money on these activities, the universe for the number of advertisers on big platforms and on the web at large is very large. ISBA represents 200 of the biggest advertisers in the UK. There are 3.5 million SMEs advertising in the UK and we know that platforms like Meta or TikTok, for example, have many millions of advertisers whose primary concern is about response to their advertising and who do not have brands that they have invested in for years and years to build their reputation.

Q192 **Emily Darlington:** Can I come back again, Phil? What you are saying is that in a newsfeed environment—your TikToks, your Instagrams, your Facebooks—there may be harmful content. Do you see that there is a consequence to brands of being near that harmful content in a newsfeed



HOUSE OF COMMONS

environment? I appreciate what you are saying—it is not going to the creator of the content—but are your members happy for TikTok to be making money off the back of the advertisers being placed near this harmful content?

Phil Smith: No, they are not. We have worked very hard with platforms to create adjacency controls to allow advertisers to have more control over where their content appears. Those controls increasingly are coming into place on the major platforms through voluntary action of the type that we have talked about before.

Q193 **Chair:** The nub of the issue, Phil Smith, is that you feel that the voluntary action is appropriate, and either it is going to change for the better or the existing situation is acceptable. To emphasise Emily's point, in addition to the implications of the revenue point, if a reputable brand, L'Oréal or Unilever, is appearing there, even if the content is not sharing that revenue, it is sharing the reputation of the brand. It is normalising the misinformation or the harmful content that is there.

I want us to be very clear. Is ISBA and the sector you represent—I know you are incredibly dependent on certain big players in the advertising sectors you talked about—calling for regulatory or legislative change? Or are you asking the Committee to wait and see how things evolve?

Phil Smith: We would certainly like to see the impact that the Online Safety Act has on content and content moderation. We are not asking for action in the arena of adjacency control for advertising at this stage. We see from the major platforms big investment in adjacency control and in the ability for advertisers to place their advertising next to suitable content.

There are other platforms that are not making similar investments, and they are not necessarily being rewarded with advertisers' money. If you put yourself in the shoes of the advertiser, it is very hard to walk away from spending with a major platform that has such a large audience, particularly among younger people, at a time when traditional media is losing its grip there. What I am saying is that voluntary action is having an effect, and while the system is not perfect, I believe we are seeing improvements there.

Q194 **Chair:** You are also telling us there is a power dynamic that limits what advertisers can call for.

Phil Smith: I would agree with that.

Chair: You agree with that; excellent. Martin and then Jon, quickly. Then we will move on.

Q195 **Martin Wrigley:** You might consider it a slightly harsh view, but would you accept that advertisers funding these platforms are actually responsible for this and should pay the consequences? Would it encourage you to get better regulation if that were the case?



HOUSE OF COMMONS

Phil Smith: It is a very hard thing for an advertiser individually to walk away from a major platform with such a huge and popular audience, which is operating legally in this country and has mass appeal, when it is striving to compete with other players who are advertising on that platform.

Collective action is clearly an illegal thing in this country. It is not something that ISBA could ever encourage its advertisers to do. Advertisers have to make their own minds up individually, but it is a competitive market.

Chair: It is illegal to the extent to which it would be anti-competitive. That is very helpful.

Q196 **Jon Pearce:** Following up on that point, Phil, if advertisers are not willing to walk away from these platforms, because that is the space where they need to advertise and to promote their products, what pressure are they placing on the platforms? Can we see it? Where is the pressure on content such as promoting anorexic lifestyles to 13-year-olds? Can we see the pressure from the advertising industry on how they have pushed back against what the platforms are doing?

Phil Smith: I can give you specific examples. If I go back to 2017 and "Brands Funding Terror", which was a headline in *The Times*, ISBA, with its members, met with YouTube and Google and asked for greater transparency, asked for greater control over their ad placements. In the following year, we saw transparency reporting coming from Google. We also saw the introduction of monetisation thresholds whereby creators were not allowed into the advertising programme to carry adverts until they proved their bona fides.

There are definitely examples in the past seven or eight years when we have called meetings with platforms. We have gone in and sought to understand what the issues are. We recognise that we are not competent in the area of content moderation, but we are certainly keen to ensure that platforms are healthy environments for their users and that our ads appear in the right places.

Q197 **Jon Pearce:** As you mentioned, ISBA and some of your members took that action back in 2017, so do you think it is time you went back to the platforms, given that there is still harmful content and it is still being connected to advertisers, and make further transparency control demands?

Phil Smith: We do, as and when the situation demands. We would have had meetings with platforms after Southport last year, for example. The other effort critical to the work we were doing was that of the Global Alliance for Responsible Media where brand advertisers were able to speak with a single voice and work with the platforms on improving transparency and controls over advertiser placement, ensuring that there was clarity for advertisers when it came to the policies in force, as well as



a degree of transparency about the effectiveness of their enforcement. That is currently the subject of legal action and I don't want to talk too much about it, but it was a major effort with global advertisers, and ISBA sat on the steering committee of that.

Q198 Chair: Phil Smith, can you confirm that for the brands you represent—perhaps you can write to the Committee—their market capitalisation or revenue is greater or lesser than that of the platforms? Do you have greater financial influence or power in the market than the platforms?

Phil Smith: Without question, we have less financial power. We could provide some numbers around that, but clearly we are a local advertising association. Many of our members are global, but none the less they are local market companies. As for the concentration of the market, we have a very small number of platforms with a very high share of an advertising market that all advertisers need to use to access, so there is definitely an imbalance.

Chair: That is very helpful.

Q199 Adam Thompson: Karen, I am going to come back to you and focus specifically on Google. We have had a longer conversation about advertising generally, but it is now very well established that Google is highly dominant in the global advertising ecosystem. It holds about 90% of the share on the sell side and somewhere between 40% and 80% on the buy side. In your opinion, what does that mean for bad practices in the advertising sector, such as the monetisation of harmful content that we have been discussing?

Dr Middleton: One of the key things to think about in terms of dominance is the inability of other players to enter the market and the fact that there is a clear gap around platforms that have safety by design. Google and the existing platforms generally have not followed safety by design principles. To the previous discussion, there are of course financial incentives for advertisers and platforms to not be regulated and to manage their own affairs, because these things are playing off safety with profit. That is the nub of what we are looking at. That power balance allows the opacity and suppression of information about how they work, along with the lack of regulation of that issue, as well as the fact that other players are not able to enter the markets.

As the previous discussion unfolded, there is of course a tension between the business interests and the safety interests. That is certainly where there is a regulatory gap where different levers should be applied to the industry, because within the advertising industry itself there is an incentive to protect economic interests which may be at odds with online safety. There are different levers, not least, first, the reputation risks for the companies, but they will not in themselves be enough to address the problem.



HOUSE OF COMMONS

Some other potential regulatory levers might be around EU legislation or regulation, the Corporate Sustainability Reporting Directive, which looks at the supply chain. At the moment, advertising is not considered in that directive, but there is a possibility for that in the future. That is something the Conscious Advertising Network is looking at around responsibility within supply chains. There is generally a problem of lack of regulation of that environment. Not to keep repeating the point, but it is incumbent on Government to find those levers. In terms of Ofcom's regulation, it is a case of waiting to see. Gaps already exist that we know about. Obviously, it is a very fast-moving environment.

As well as regulation coming on board, there are current developments with some of the platforms, including Meta's move to roll back terms of service, which a report by the CCDH estimated would mean that 97% of the content they previously removed will not be removed. That definitely exposes the problem of content moderation, because it relies on the platform having particular terms of service and removing content once it is already out there. The regulatory approach would make it more incumbent on any of the services to ensure that they have the right kind of safety by design.

Q200 Adam Thompson: You talked repeatedly about safety by design—I think that has come through very clearly—and the difficulty for competition entering the market. Do you think that more competition will make a difference to the spread of harmful content, or is the business model itself to blame? You also said that the Government should be doing more. We asked Phil a few minutes ago what he thought the Government should be doing. What are your thoughts on that?

Dr Middleton: What should the Government be doing? I am sorry. What was the previous question?

Q201 Adam Thompson: The first question was: do you think that more competition would make a difference to the spread of harmful content?

Dr Middleton: In terms of competition, it is very difficult. There are barriers to entry and there is market dominance, so I don't think we could take a competitive position and say that that would by natural market forces reduce the problem, nor will the reputational risk necessarily be enough to reduce the problem. Companies themselves can advance their own contractual arrangements; they can do further risk assessment of the algorithms and how the prioritisation of certain content that is monetised may have an effect on their own reputation, but that in itself is not, I think, enough to drive progress. It is incumbent on authorities in this space, which would be the ASA, the Information Commissioner's Office, Ofcom and potentially the Financial Conduct Authority, as I mentioned earlier, to consider this. Obviously, we need to consider the economic interests of the companies, but just relying on that is not going to deliver safety by design outcomes.

Q202 Adam Thompson: What are your thoughts on what more Government



HOUSE OF COMMONS

could be doing than they are currently doing?

Dr Middleton: To date, Ofcom has not mentioned anything about misinformation. There is an online advertising taskforce, but it has not talked about that issue, so that could be a development. To my knowledge, alongside the other bodies voicing it as a problem to be addressed, it will not be addressed overnight, but it is within the responsibility of the industry.

Q203 **Adam Thompson:** You said that companies could be doing more as well. Given the dominance Google has in this space, do you think it is doing enough to avoid monetising harmful content?

Dr Middleton: From where I see things, it is not. I know that in one of the previous meetings in this inquiry, they talked about an API that they are releasing. That is a positive step, but Google commands a high proportion of this marketplace. It devotes resources to content moderation. However, I do not think there is enough in terms of safety by design.

Q204 **Adam Thompson:** Phil, looking at previous interventions, such as the online advertising programme, the Online Advertising Taskforce and the intermediary and platform principles pilot, where do you think they have fallen short on tackling harms in the digital advertising market?

Phil Smith: Let me start by saying that I think all of those initiatives have a different end in mind. If the issue we are really talking about here is harmful content, we have obviously talked a lot about the Online Safety Act, so if Government believe that the Online Safety Act needs to go further and other harms need to be included within its scope, clearly that is something they would need to consult on and ISBA would be really happy to help in that regard. Clearly, misinformation does not currently fall within the scope of that.

When it comes to existing initiatives, the online advertising programme is intended to address harmful advertising, not harmful content, particularly ads that might be harmful to children. We have been deeply engaged in that work. We lead two of the working groups. I chair the one that aims to create benchmark data on how effectively advertisers of age-restricted products are targeting their ads when they follow the codes for adult audiences, where we believe we are actually doing a better job than legislators might think. The intermediary and platforms protocol from the ASA is intended to bind platforms more closely into the ASA system where, as Karen said, the remit is about the content of ads as we move to a more automated proactive system, to ensure that advertisers are aware of their obligations to the ASA. All the indications from the pilot are that it has been successful, but its remit is different and limited compared with the scope of this inquiry, which in my view is much more about content and how the business models drive that content.

Q205 **Adam Thompson:** You mentioned the UK online advertising programme. From your perspective, what is the current status of that? Where is that



work at the moment, and where do you see it going in the next months and years?

Phil Smith: We will see whether or not it has succeeded when we get to the autumn, and each of the working groups will be able to demonstrate that it has a plan in place to meet its goals. The progress I have seen so far through the status updates is promising and encouraging.

Q206 **Chair:** To clarify, in the previous discussion about the online advertising programme and taskforce and intermediary and platform principles, you emphasised that that was more about the content than the process. You also mentioned the Global Alliance for Responsible Media, which was the voluntary cross-industry initiative that defined harmful online content, where the advertising sector tried to play a role in reducing the possibility of advertising supporting harmful content. In August 2024, Musk sued GARM. You implied that legal proceedings were ongoing. Does GARM still exist? What is the status of that? Is it acceptable that an alliance for responsible media can be hounded out of existence by the threat of being sued?

Phil Smith: I decline to comment on the last part of that question, but to clarify the facts, in August 2024 GARM ceased its existence; it was withdrawn by the World Federation of Advertisers. That lawsuit is ongoing, and a further tranche of advertisers has been added to it by X. It will take some time for it to run its course through the courts in Texas.

Q207 **Chair:** That suit is happening in Texas, but GARM's activities have been suspended or closed globally.

Phil Smith: Globally, they have.

Q208 **Chair:** Responsibility by responsible media has not been supported through self-regulation or industry groups in that particular example.

Phil Smith: You wouldn't say there were a lot of participants in GARM on both sides of the fence who were very positive about what was being done, and one was less than happy.

Chair: The Committee notes that Elon Musk, who sued GARM, declined to appear before this Committee and has indicated his view that this Committee's proceedings are contrary to freedom of speech. He seems to be somewhat contrary to freedom of speech himself.

Q209 **Dr Sullivan:** We have touched a little on self-regulation. Karen, you reported on maybe the need for stricter codes. What can we be doing to stop the practice of harmful digital marketing?

Dr Middleton: I would answer that question by saying that the first step is clearly bringing the problem to light, which is obviously what we are trying to do here, and that within those bodies the problem is considered. To date, the ASA, for example, has considered the content of advertising, but it is the Advertising Standards Authority. That does not limit it to having only that remit. In looking at standards for advertising, surely if



there is a large-scale problem, such as the one we have here, it should be within the remit of a regulatory body such as the Advertising Standards Authority.

One of the recent codes the Advertising Standards Authority has added is on advertising making environmental claims, which is reasonably broad and not as straightforward as what it has been looking at to date, which is harmful high-fat, salt and sugar food advertising, stereotyping and that kind of thing. Making environmental claims that are not justifiable by the organisation and would represent greenwash requires looking at the whole business practice of that organisation and making a judgment on whether it is greenwash or not. There is a little bit of broadening out, but if the body exists in order to assist the industry to practise responsibly, maintain standards and not be associated with harm, I do not see why it should not bring something in.

The whole problem is very complicated, opaque and difficult to pin down, which is why I think there would need to be multi-level consideration of the problem and its output. For example, the Advertising Standards Authority could issue guidance that advertisers should be aware of this potential problem and take steps to audit their supply chain; they should take steps to use the tools Phil was talking about; they could perhaps do risk assessments and seek contractual agreements, if possible, but there is obviously a recognition that there are larger and smaller organisations with greater or lesser resource to do those things. I guess the ASA could issue guidance on how advertisers, platforms and websites within the supply chain should be aware and consider addressing the issue themselves.

Q210 Dr Sullivan: At the moment, it feels like the wild west on the worldwide web. We can issue guidance; we can say, "Here's best practice," but that means nothing unless there are actual teeth and enforcement. Would you propose a regulatory body that includes a kind of enforcement angle? What would be a disincentive for bad practice?

Dr Middleton: Ofcom fines. We are waiting to see how those would be implemented. This is an area that draws a lot of passion and there has been a huge amount of positive work by Ofcom, but rightly there is criticism about whether it goes far enough. There is a kind of anticipation about whether the disincentive in fines will be taken up, but from the outset up to 10% of global income could represent a fine.

Q211 Chair: To clarify, that is for Ofcom's powers with regard to fraudulent advertising and harmful illegal content.

Dr Middleton: Yes.

Q212 Chair: Their powers are not with regard to the process of advertising or misinformation, which are I think the main concerns of this Committee.

Dr Middleton: Absolutely. I am sorry not to be clear on that. What I mean is something similar to that where Ofcom has powers to impose



HOUSE OF COMMONS

finer. In terms of having teeth, Ofcom would be the only regulator unless there was a new kind of regulator, which I think is very unlikely.

Q213 **Dr Sullivan:** Phil, do you have any comments on those points?

Phil Smith: The ASA is moving rapidly to an automated proactive system that operates at scale. Its remit, which is funded by brand advertisers, is the regulation of legal advertising content by good actors trying to do the right thing. It is not enforcement against illegal activity.

My other point is that there is risk of regulatory overlap. We already have Ofcom operating in this space, and in this Committee we are talking about the interplay between content and advertising. For our part, we have been pretty clear that content is not within the remit of advertisers to seek to regulate. If there is more power that Ofcom needs in this environment, it is something that we think Government would need to consider.

Chair: I think we have established that there is not a regulator for the process of advertising, if you like. The Committee has had evidence that it is the process of advertising which, as well as driving huge revenues and profit, drives misinformation. That is a regulatory gap.

Q214 **Emily Darlington:** I refer to something that has come up quite a lot. We had some bereaved families of teenagers at our previous session. Dr Karen Middleton, in cases like that of Molly Russell it is very clear from the coroner's report that he believes that exposure to harmful online content contributed to her death. Her family very much wants to see the messages and data for children who have died. The coroner said it was quite difficult to get the information in his inquest. Should this data be shared with bereaved families, and should it be made easier for coroners to access the material?

Dr Middleton: In answer to your first question, I do not think that as a matter of course families should be given that data. That could be very damaging and an invasion of privacy, but the data should be available to the coroner if required. That should be a decision for the coroner to take. It is not my particular area of expertise, but it is definitely something I keep up with in the research. I believe it was proposed, but at the moment there is no requirement for the platforms to make that data available. I think it should be made available to the coroner.

Emily Darlington: The Government have moved, and it is in the Data (Use and Access) Bill currently going through the Houses. I guess it takes us back to the question Martin posed. If these platforms are publishers, who owns the content that is created on them and who ultimately has responsibility and accountability for the consequences, particularly when young people are exposed to that material? Where does the liability lie?

Chair: On your point about the invasion of privacy, unfortunately we are dealing specifically with a case involving children who have lost their lives and the families are desperately trying to understand what happened. In



the offline world, as it were, the families would inherit access to any diaries or notes. It seems that this is a case where the offline and online worlds diverge.

Q215 Emily Darlington: It is quite important for us to understand the impact this is having on particular behaviours, when some of those behaviours lead to them taking their own lives. How are we as a Government to assess that if we do not have the data that shows the influence of it one way or the other?

Dr Middleton: Indirectly, it raises the question of terms of service and GDPR-related issues. The individual, albeit that they are a child, still has some agency over their own data. A related issue, which we obviously do not have time to go into, is that the terms of service are themselves difficult for individuals to access. They are in some ways rendered meaningless because they are very inconvenient and hidden. If you were to ask for it to be commonplace for everybody to make a decision about whether they would want that information to be made available to their family in the event of their death, it would just be under a long list of terms of service that currently are quite difficult to navigate.

Chair: Thank you very much. I am sure that is a point the Committee will return to. We have to leave it there. Dr Karen Middleton and Phil Smith, thank you. We have found your evidence informative and enlightening, and thank you for spending the time with us this morning.

Examination of witnesses

Witnesses: Dr Rahman and Lyric Jain.

Q216 Chair: Welcome to the second panel in the Select Committee's session this morning on social media algorithms and harmful content. We will be looking now at content moderation and online safety. I welcome our two witnesses. I will start with a question to Lyric Jain and give you the opportunity to introduce yourself, and then, Dr Eirliani Abdul Rahman, you can introduce yourself later.

This Committee is particularly looking at the causes and aftermath of the riots over the summer. As Ofcom said, there was a clear connection between online activity and the violent disorder on the streets. The ways in which social media were implicated included a post on LinkedIn claiming that the alleged attacker was a migrant, which was viewed more than 2 million times despite being taken down, and an X user sharing a false claim of the attacker's name, which was used as part of TikTok's "others searched for" feature promoting it across TikTok. Lyric Jain, do social media platforms have adequate measures to deal with misleading and harmful content in times of crisis?

Lyric Jain: It is a pleasure to be here. Thank you to everyone here for the opportunity to provide evidence. I am Lyric. I am the founder and CEO of Logically. We are an artificial intelligence company very much founded to build technologies to identify information operations in the



online landscape, and we have been working with Governments and platforms over the last four or five years to do exactly that.

We were quite active last year amidst the events in Southport. There are a number of things to take into account when looking at how effective platforms' measures were. The first point when speaking about the very large platforms is that often typical discussions focus a lot of the debate on what they should be doing. We saw that there was a lot of activity on not necessarily the very large online platforms but emerging large platforms. There is a significant difference between the measures and capabilities that very large platforms have and even large platforms have. The distinction between the two could be platforms with tens of millions of users and platforms with millions of users. There is a significant gulf in capability between what those two sets of platforms can do.

Q217 **Chair:** Could you name some of those platforms?

Lytic Jain: A platform that typically has not been in the VLOPs category would include Telegram. It has recently been reclassified into the VLOPs category. A lot of the activity related to Southport originated on Telegram, particularly foreign interference; there was significant foreign interference on Telegram as well as X.

To really answer the question, it is worth considering how this space and the capabilities have evolved over the last five years, or even eight years if we are taking a longer view. Typically, the measures that platforms have in place to tackle misinformation have been very distinct and largely separated from how they counter disinformation or tackle content likely to incite violence. The ways in which those two programmes work, even in the very large platforms such as Meta and TikTok, are largely quite disconnected, to the point where they have very separate programmes, the teams have very limited collaboration and there is very little collaboration even between platforms when these instances flare up.

One thing to note is the change in how especially the foreign element has operated over the last few years. If we zoom back to 2016 and the Russian information operations in particular that operated back then, the modus operandi was content farms. It was how to initiate messaging in a number of countries to achieve whatever the objectives of the Russian Government were. Since 2020, that has had a pretty significant shift. They very rarely originate new messaging. They amplify messaging that already exists within domestic communities, and that makes it very hard to know where a foreign influence operation starts, where organic domestic participation around it begins, and, equally, what enforcement of accounts and content that have been engaging in that campaign should look like and how it should be different for domestic accounts versus foreign accounts. That is very much one of the underlying challenges that we saw in Southport.

Finally, there was a lot of amplification synthetically for what we saw. Everyone can agree that we should have stringent freedom of expression



protections for the population of the UK and citizens of the UK. However, our view is that bots should not enjoy that privilege. It was certainly the case on X that bots played a significant role in amplifying some of that activity.

Q218 Chair: What do you think Government should do to address the challenges and the deep concerns that you have outlined? In terms of our domestic security, I think you said the platforms do not have adequate measures. Do the Government and our security services have adequate powers to address that?

Lyric Jain: It is somewhat unclear at this stage because the OSA was not in full force as of last year. The advisory board of Ofcom is still being set up, so it still remains to be seen how a lot of these measures will be effective. There is great hope that existing legislation would help reduce the impact of such activity in the future. The ground that is perhaps less covered through Government intervention and action at the moment is looking at proactive monitoring. The legislation makes some note of what Government are currently doing to proactively monitor for these incidents and the requirements they place on platforms to proactively monitor for these incidents. Platforms are currently motivated to do so on a periodic basis—a monthly, quarterly or annual basis—based on whatever the country's reporting requirements are. They are not particularly motivated based on risk. They are not motivated based on risk that day, that hour or that week on public safety, public health and national security. There could be a more risk-based approach to intervention and the approach with which we assess platforms on how effective their measures are.

On the Government front, the biggest thing in the Southport case that would have helped is more disclosure and more information. The vacuum of information in the initial hours did not help. If both through local policing and through selective declassification of intelligence there could have been early sharing of information in the public domain, that could have perhaps been an effective intervention by different parts of Government.

Q219 Chair: I am not only talking about effective intervention at that point, but about Government powers and regulation. You said that we look to the Online Safety Act to provide more powers, but the social media companies told us that they would not have done anything differently had the Online Safety Act been in force. Did we know which messages Russia was amplifying at that time, and did we have the powers to address them?

Lyric Jain: There was great suspicion at that time. This remains one of the challenges in attributing foreign interference. It is quite simple to do so retrospectively after we have all the fingerprints and we can see how campaigns evolved, but not in real time as campaigns are under way online. We did not even have a lot of similar capabilities operating in Government. It took a few hours before we were able to make that determination, and by that point in time the narrative was already out of



control. What I would call for from a capability perspective is more investment in real-time capability in identifying things on an hourly and minute-to-minute basis as opposed to moving with timescales of days in mind.

Q220 Emily Darlington: I want to talk about your point on anonymised sources, whether they be agents of another country or people using the anonymity of profiles. Do you think that is something that Government should look at in terms of ensuring that there is capability to know exactly who is running what profiles? It might be linked to the Government's efforts around digital identity.

Lyric Jain: It is a challenging intervention because we want the ability for people to speak anonymously and freely online, and adding additional barriers to that could potentially demotivate people—say, whistleblowers and others—from making such public disclosures. In principle, such interventions would help, but the balance between that and people's ability to express things freely, even when they might be under certain threats, needs to be considered while framing that.

Q221 Chair: Thank you, Lyric Jain. Let me come to you now, Dr Eirliani Abdul Rahman. Meta told the Committee that safety is its No. 1 priority, while TikTok labelled safety its north star since day one. Do you agree with those claims?

Dr Rahman: I read through the testimony from the previous session, and I would like to restrict my comments to X and Meta.

Q222 Chair: Could you introduce yourself first?

Dr Rahman: I am Dr Eirliani Abdul Rahman. I am a senior fellow at Georgetown and consider myself a scholar petitioner. I would like to restrict my comments to Meta and X. At the time of Elon Musk's takeover of Twitter, I was on the Twitter trust and safety council. I resigned in protest against the meteoric rise in hate speech after a month in. Following Mark Zuckerberg's announcements in January this year, I do not believe that a tech billionaire's focus on safety is a No. 1 priority. I want to give a quick overview of the last two years in terms of content moderation and developments in online safety. From November 2022—

Q223 Chair: Could you lean in a bit to the mic?

Dr Rahman: Sorry, is that better now?

Chair: Yes.

Dr Rahman: From November 2022 until November 2023, Meta, X and YouTube de-prioritised content moderation and rolled back policies that dampened hate and harassment, disinformation and misinformation on their platforms. In total, they removed more than 40,000 workers who were working on trust and safety and platform integrity. I am going to zoom in on X because that is where I was. Those cuts reflected the company's priorities and values. Musk fired more than 7,000 people,



HOUSE OF COMMONS

which comprised 82% of the workforce. It had serious impact on the core capabilities of the platform. There was a huge rise in hate speech and disinformation. There was a study by the Institute for Strategic Dialogue that showed that English language antisemitic tweets more than doubled in the period following the takeover.

I want to highlight that whatever Musk does is almost like a playbook for the other companies. In November 2022, he reinstated the then former US President Donald Trump's account. By doing so, he basically normalised, by not taking such actors' incendiary speech into account, and privileged such speech. In January 2023, Meta followed suit, and then in March 2023 YouTube. In the same period, X also reinstated previously suspended accounts that belonged to white supremacists, misogynists and those who promote hate speech. Unfortunately, one of them was Tommy Robinson—full name Stephen Yaxley-Lennon. That is just on the platforms and capabilities.

I want to give two examples regarding safety on these platforms. One is Meta, and it has often been cited, but it bears repeating: Myanmar. In 2021, Amnesty International did an investigation that found that Facebook substantially contributed to the atrocities perpetrated against the Rohingya. From 2013 to 2017, there were multiple calls from civil society about the failures of Meta in that country. Meta chose not to heed those warnings.

In the first week following the Israel-Hamas war, there was analysis by NewsGuard as an actor interested in safety, and it found that X had become a super-spreader of misinformation regarding the conflict. This was boosted mainly by the blue checkmarks. In those days, it was more to authenticate a user, but these days you can just buy it, as you all know. These users on X account for 74% of the most viral falsehoods relating to the conflict. In the first week alone, there were 1.3 million engagements with those posts and more than 100 million views. NewsGuard found that the false narratives tend to go viral first on X and then spread—migrate—to other platforms, including TikTok and Instagram. I disagree that it is safe. We can also go into further details, including what Fernandez mentioned about the global transparency report, because I have more numbers on that as well.

Q224 Chair: Thank you very much for that contribution. You emphasise the consequences of the spread of misinformation in the past and abroad. Specifically on the Rohingya massacres, has Meta, in your view, been held to account for that in any sense?

Dr Rahman: No, it has not. There is an ongoing legal suit by defendants in that space, and Meta has chosen not to respond, as far as I am aware.

Q225 Paul Waugh: Thanks for that, Dr Rahman. I want to check something that X told us. Under their new rules, they say they restrict the reach of any posts that will contradict their own alleged rules on violent speech, harassment and abuse. They say they restrict the reach, but they say



that they do not remove it. Is that a distinction that is not worth the candle? In other words, is there any point in them having rules at all if they do not remove content? Is it justifiable in any way for them to say they restrict the reach of that content?

Dr Rahman: It is really interesting because a lot of the oral proceedings that I have gone through in January and February focus on content. That really protects the companies under section 230 of the US Communications Decency Act. For me, it is important to focus also on the actors and their behaviours. I want to highlight the work of Professor Camille François from Columbia University. She has highlighted the ABC framework on viral deception. She calls it three vectors: A for actors, B for behaviours and C for content. As long as you talk about content, the platforms can always say, "Section 230 doesn't apply to us." In different jurisdictions, such as the OSA, although I am not an expert on it, there are ways that we can improve on it.

It is important to understand that amplifiers—in this case, highly networked accounts—play a significant role. For example, Tommy Robinson says certain things on different platforms. Those are amplified by what I call swarmers, people who contribute to the swarm. We talked earlier about the identity of those people. They do not have an identify; they are trying to be anonymous. The amplifiers play a very big role in how things get inflamed. An important thing that we need to look at is the role of such actors. In the particular case of Musk, because he owns the platform, he has an outsized influence on what happens. With the Gaza war, a study done by the University of Washington has shown that Musk engaged with at least six of the seven top propagators of disinformation in October 2023. Because he did so, they took off. We need to have a better understanding of what it means for amplifiers to get involved, and then what it means when the swarms take off and how it impacts what happens in terms of offline violence.

Q226 **Paul Waugh:** Is there any evidence that X's claim that they are restricting the reach of this content is actually happening?

Dr Rahman: It is very hard to say because access to the API is prohibitively expensive. One of the things the OSA can do would be to have mandatory data access for researchers to the regime that there was before, when we were able to actually say, "Okay, this is what happens," and then we can check on what they are saying. A lot of it right now is self-reporting. They say, "We are doing this," but we cannot really track unless we have that high-level access. I cannot really comment.

Q227 **Chair:** Thank you very much for that. You highlighted there the role of Elon Musk in amplifying misinformation during the early days of the Gaza conflict. Are you aware of any way in which that misinformation spread or his actions are being held to account?

Dr Rahman: Not that I am aware of, no.

Q228 **Emily Darlington:** Can I give another example? I won't read the tweets.



HOUSE OF COMMONS

When Tommy Robinson was readmitted to X, there was a particular amplification by Elon Musk, particularly the documentary Robinson is currently in prison for in my constituency, so I get a lot of the swarm for that. It would have detracted from the prosecution of people who were going on trial for grooming gangs. That is what he is in prison for; so that he could not do that. Given that it involves a legal case and given that Elon Musk was amplifying messages that could have led to the breakdown of the trial of those child groomers, which would have been appalling, do you think the platform and the individuals should be held to account? What would the legal framework need to look like in order to do that?

Dr Rahman: It is currently insufficient. If you look at hateful conduct and abuse and harassment, some of the same actors are going against the terms of services. It is quite clear and it is quite simple. If you have an amplifier who is also an owner of a platform and they are doing certain things where it is hateful conduct and/or amplifying harassment, I do not think the current legal regime as it stands in the US would protect any of the users. I do not know whether you have a copy of the note that I sent. This is not new; we all know about dogpiling and brigading, but, shockingly, it has not to my knowledge been quantified before. I quantified what happened.

Basically, you have someone who is highly networked and an amplifier. For example, as a Member of Parliament, Emily Darlington, you may be a target. He puts you in front of a new audience, and you have a sudden jump. You are just doing your normal work. You tweet about something, and he does not like your tweet. He just takes it up and he puts it there in front of a new audience, and you have a sudden jump on this particular tweet, a stochastic increase. All you need to see is not so much what is in the content of what I call the swarm, because you see a sudden rise, but the pattern. When somebody like him posts and you see the jump, that is all you need to know for a platform. You see this happens, the amplifier posts, and the swarm ensues. We can ask the social media platforms to establish thresholds for that form of harassment. I call it indirect swarming because you catalyse a swarm. Most of them are against you. You have followers who will try to support you, but in doing so they still contribute to the swarm, so it is inevitable. In my case, I received death threats and doxxing threats alike.

They were quite smart about it. They did not say they wanted to see me killed, but they said in graphic ways how they want to see me die, so I could not take it to the FBI. I had to keep checking my timeline in order to make sure there were no imminent death threats. For the other two women who were alongside me, it was too much for them. I was the one who was checking our timeline five times in that moment. The kind of jump that I am talking about here in our case was 200 tweets every minute over a 24-hour period. Social media companies know that that is a sudden jump that is not typical of what was happening before. They can establish those thresholds. Ofcom can have monthly or quarterly



HOUSE OF COMMONS

assessments with the companies to see what that looks like in terms of establishing thresholds and levels.

What is most important is that the OSA tries to establish more user controls and give agency back to the user. With that approach, if companies allow us to have a button that says, "I'm being mass harassed, I'm being dogpiled," or whatever it is, you press the button and you stop all engagement. On Twitter right now—in my case, I am off it—you can only block five people at a time. That is an impossibility. They can still like it and retweet it, which still acts on the algorithm, still acts on your visibility and still allows the virality to continue. A lot of what we have been talking about so far in January and February is content. We need to talk about the speed, the amplification, the volume and how it is done in terms of posts and engagement. That is what is really pertinent. The companies know that. They are not going to tell you about it.

Chair: Thank you very much. That is very interesting. Adam, you want to come in on that point.

Q229 **Adam Thompson:** Yes; thank you very much, Chair. Thank you very much, Eirliani for your explanation of dogpiling and that whole process. I know from my own social media that every time I post about education policy specifically I have one or two people who will take that post and share it in a group, and I repeatedly get dogpiled by the same group of individuals who specifically oppose our particular education policy, so thank you very much for that.

You talk about how you might be able to resolve some of those issues by including an "I'm being dogpiled" button. This is the safety by design methodology we heard on our previous panel from Dr Karen Middleton. I was going to ask if you think that it is enough to rely on content moderation. I think you have made it pretty clear that it probably isn't. Could you elaborate on some of those thoughts about implementing safety by design and what it might look like in practice?

Dr Rahman: It is a great question. A lot of the focus in the debate in society in general has been on the supply side. The narrative has been led by the big tech companies themselves. They co-opt the language. They use language like "empowering" and "democracy" in order to de-prioritise content moderation and push the onus, the burden, on to us. As the Committee knows, it is crowdsourcing. For me, that is the wrong approach.

This inquiry is a timely reminder of what is important. It is not just about content moderation per se as a stand-alone thing. We need also to interrogate critically the entire system, not just content moderation but what the platform was set up for, what problem it was meant to solve, what communities they are fostering online, their norms, their practices, and whether they are inadvertently, or maybe even not inadvertently, entrenching online harms by the structure that they have built.



A more pertinent question is: what do we want as users? To your question, I like that very much, because we know this. My field of expertise is network harassment. That means somebody who is an amplifier puts you to a new audience and you get harassed. It is disproportionately affecting people of colour, women and people from historically marginalised communities. There is just no way around it. That is the truth. Are their voices included when we talk about, say, safety by design? It is not the case. We may have stakeholder engagement, but are they being significantly included when you talk about processes and outcomes? Musk again has led the playbook in this regard because he basically got rid of the entire council. I left the council, and then he fired the rest who remained four days later. Other companies followed suit because they see that it can be done. Child safety is not a thing any more. A lot of it has been outsourced. We need to understand what it means when that capability is not there any more.

I want to highlight a UNESCO report from 2023 that I find quite disturbing. It is on tech-facilitated gender-based violence. What it argues is that it is not just about online harassment; it is really a test bed for new and novel forms of emerging harms. That is the problem with technology. If we talk about legislation, we are always playing catch-up. People who are doing the trolling or doing this kind of work know what they are doing, and they are using us—people of colour and women—as a test bed before this form of harm goes out to the mainstream. It is actually quite dangerous. What is the question again?

Q230 Adam Thompson: How do you think you could implement these? What sort of safety by design protocols might be implemented in practice? What could they do?

Dr Rahman: Meaningful engagement, so it is not just a tick in a box. It has to be sitting at the table and then co-creating something together. I do not think currently with technology and big tech the ethos stands, because it is really moving fast and breaking things. If we are talking about safety by design, it is slow and it is deliberate. It is talking about the things that I just mentioned where we have a button, which goes against the business principles, because you want more engagement. You want more toxic engagement, to be more precise. If you switch it off, that will impact the business models. Frances Haugen said it very clearly. There were ways to look at things that were not based on content, and it would basically turn it down by a few percentage points.

Facebook was reluctant to do it. They had very clear things where in 2019 they were trying to figure out the algorithms. This is from the “Facebook Files”, a series of stories published in the *Wall Street Journal*, based on leaked documents by Frances Haugen. Even as far back as 2019, Facebook was really worried about the algorithm and what it could mean. It was not until 6 January 2021 that they realised that they had to change the way the algorithm was working because it was causing more harms. We have talked about inauthentic co-ordinated behaviours like



bots and state actors. Facebook was looking at the time at authentic actors too—people like the Patriot party and all those things—and they knew they had no policy around it. When Meta say that they have all these announcements and would not take proactive action on certain forms of harms, what does it mean? Does it mean this form of harm? Does it mean incitement to violence? Which bit? I am not clear.

Q231 Adam Thompson: Thank you for that. That is incredibly clear. Two final questions from me. First, I think you have answered this already, but to clarify, do you think it would be technically possible to implement the types of changes that we are talking about in these platforms? If we assume that it is against the business interests of the companies to implement these kinds of tools, how can we force them to do it?

Dr Rahman: We must focus on threshold ratios and making it user centred. You can say, “We have this form of new indirect swarming where threshold levels are like this.” I presented this work at the Stanford Trust and Safety Conference and at the TrustCon regional meeting, as well as the summit. It really resonates with engineers because they think it is eminently doable. It is also a user tool that can be done to prevent virality. Usually, we think in industry that it could be three hours or it could be quicker or longer, but it stops it from taking off. That is important because it is user centred. It is compliant with section 230 of the US Communications Decency Act. It preserves the first amendment right to free speech because the user decides. It does not matter whether it is bipartisan or whatever it is. It is just virality. If it is negative virality of any form, the user decides, and that is it. It is completely doable. I have spoken to people in big tech for their insight, and they think it can be done.

Q232 Adam Thompson: If we work under the assumption that implementing these kinds of controls would inherently reduce engagement with platforms, how do we tell them they have to do it?

Dr Rahman: It is like the tobacco industry or any other industries in the past. We need to have a public health lens on big tech. Right now, it is a bit like the wild west. They would disavow this, but it is true; we let them dictate the narrative on how it can be done, when actually there are ways around it and there are ways to do it. Let’s take a public health lens approach. We are always talking about the intention of the harasser and the intention of the perpetrator, which is hard to adjudicate. If we talk about the impact on society and the impact on targets and victims, that is clearer. It is a bit like putting on a seatbelt and going out to drive. Your intention was not to drive over somebody, but you did, and that person died. We are talking about what happened to that person. It is a very simple thing. Seatbelts and smoking—there are precedents. We need to do the same in this industry.

Chair: Thank you very much. We know that the Online Safety Act gives Ofcom the ability to look at platforms’ algorithms. It also gives platforms the duty to mitigate certain risks. We are not clear whether



HOUSE OF COMMONS

misinformation is one of them. Between that duty and that ability, there would certainly be the opportunity to have transparency on whether those mitigations and dial-down controls are in place. That is certainly something that we can raise with Ofcom when they come to see us. Thank you so much for that.

Q233 Paul Waugh: Dr Rahman, we had the representatives of X and Meta before us the other week. They were softly spoken, but what they said was still shocking. When I put to them in particular the leaked guidance about Meta, particularly the fact that statements that could now be allowed are “Black people are more violent than whites” and “Jews are flat-out greedier than Christians,” the response from Mr Yiu was that, “some conversations, while challenging, should have a space to be discussed.” You used to sit on, effectively, Twitter’s regulatory arm of the safety issue. How dangerous a shift has that been?

Dr Rahman: It is a very big shift. I also want to mention that Meta submitted written evidence for this inquiry on 18 December 2024, and just 20 days later Mark Zuckerberg did a complete turnaround and said that he does not want to do proactive enforcement of certain forms of harmful content—what you are alluding to. I find it really shocking because it does not make the platform safe for anyone. We need clarity on what that means because it is not clear at all what will happen now. He talked about community notes. I can talk about what that means in terms of efficacy. It is not efficacious at all. I worry about users and how we can make sure that we protect users’ safety.

Q234 Paul Waugh: The representative of X said that he felt that X was now safer under Elon Musk. What is your reaction to that?

Dr Rahman: I really wanted to prepare for this question. It is not safe at all. I had a look at the global transparency report that he mentioned, which was for the first half of 2024. It marks the first time that X has published such a report under the new regime under Elon Musk. The previous such report was for the second half of 2021. For that one, I just want to mention the numbers. For 2021, 11.6 million accounts were flagged for violations, of which 4.3 million accounts were acted on and 1.3 million were suspended. In the latest report that Fernandez mentioned, there were over 224 million reports of accounts of individual content, and yet despite the huge rise the number of suspensions only rose to 5.3 million. This is comparing apples and oranges because Elon Musk had a whole policy shift with his rollback on hate speech and allowing misgendering on the platform. Let’s talk about hateful conduct. In 2021, there was action taken against 1 million accounts for posting hateful material. In this latest report, there were only 2,361 accounts addressed for hateful conduct. How many reports were there for hateful conduct alone? There were 66.9 million. The numbers themselves, even if you talk about takedown rates, are shocking. I will give you more numbers now.

Chair: Sorry, we need to move on. Maybe you can write to the



Committee.

Q235 Paul Waugh: One of the things that we are interested in is what the UK Government can do to combat this. Is there anything extra under the Online Safety Act that we should be looking at as a Government and as a Parliament to stem that surge, as you call it, in hate?

Dr Rahman: I am not an expert on the OSA, but one of the things that is not mandatory right now is data access. That would be a really good way to look at what X is actually saying. A lot of it is really, as was said in previous sessions, marking their own homework. If we have access to the data, it is much easier to do some of the things that I mentioned, even the threshold ratios. We can have case studies. We can show longitudinal work, even in terms of what X is saying about safety or not. A lot of these things can be proven and done by academic researchers themselves, but we need data access, and that would change a lot of things.

Q236 Paul Waugh: Maybe legislate to force them to be more transparent.

Dr Rahman: That's right.

Chair: There are options that we need to look into. Thank you so much, Dr Abdul Rahman. Lyric Jain has not spoken for a little while, but I know Lauren has a question for him.

Q237 Dr Sullivan: Thank you very much. Yes, I have a question for you, Lyric. What do you think are the most effective ways of preventing the spread of misleading information? Is it an independent third-party checker? Is it a crowdsourcing model? Is it a combination? What do you think we need to be doing to stop the spread of misleading information?

Lyric Jain: There have been a lot of learnings over the last six or seven years since Meta initially rolled out its 3PFC programme. At the moment, their approach seems to be community notes. There are some really good aspects about community notes: more people can participate through crowdsourcing; people with greater local context or subject matter expertise may be able to contribute to assessments; and it is certainly a lot more economical. The X implementation of community notes promotes conflict in a good way—a conflict of ideas. There are certainly some good elements to community notes.

What is somewhat unsatisfying about community notes from a user experience perspective and from a safety perspective is that they do not always lead to a definitive conclusion and they are not always timely. If we look at the work that 120-odd accredited fact-checking organisations around the world have done as part of Meta's programme or TikTok's programme—full disclosure: we happen to be part of those programmes—it has been work that has been powerful in helping to downrank certain types of harmful content. However, a lot of it has not had a sufficient level of technology support to really be motivated towards risk. Over the last six, seven or eight years that these programmes have been operational, a lot of the fact checks were really



impactful. A lot of the fact checks that were conducted by organisations were on quite a few frivolous topics. They were quite repetitive in some cases. Looking at it from a platforms perspective, there is certainly room for innovation and change.

What we advocate is a hybrid approach. When it comes to the riskiest pieces of communication and misleading content, which could have the most significant implications for public health, public safety and national security, we need to take a risk-based approach where those types of narratives and those types of misleading content are prioritised through feeds for professionals accredited to assess, whereas for the wide swathe of online communications something like a community note-style intervention might be helpful. Even in the last two months, despite some of the challenging news from platforms and their shifts in policies, there has been some appetite from them to take on such approaches. What would help perhaps in the UK context is mandating platforms to have something like a 3PFC programme.

Q238 **Chair:** Something like?

Lyric Jain: Something like Europe does currently right now. Under the DSA, platforms are required to report on whether they have any third-party fact-checking programmes where professionals are used to assess for misinformation risk and so on. A similar approach in the UK would be quite welcome. We work with Meta, and we will continue to work with them through this year. It appears that the community notes programme will be rolled out globally on Meta and other platforms some time next year. It would be helpful if UK legislation were in a place where platforms were mandated to do something in that space. At the moment, there is a gap. They are mandated to do something in Europe, but not necessarily in the UK.

Q239 **Dr Sullivan:** It is a bit of both that you are promoting there. Regarding the third-party professional for mental health, public health, national security and those sorts of things, given the huge numbers of posts that there are in different community groups, scrolling, it is going to cost a lot, isn't it, for those companies to do that? Do you think that they may not wish to do it for those reasons?

Lyric Jain: There are definitely economic reasons not to do so, which is why prioritisation is important. Meta invests something like \$40 million or \$50 million a year specifically on their 3PFC programme, which is significant, but not that significant when you consider that their total trust and safety budget is about \$2 billion. It is not a category that a lot of investment goes towards, and Meta is the leading example of investment there. Through technology and through a lot of the progress that has been made in AI, there can be a lot more efficiency driven into the expert verification process. The challenge over the last six, seven or eight years since these programmes have been alive has been how we get to the stuff that is most harmful in a timely manner. That is the area where technology can make the greatest contribution to get a greater



impact from the work that fact-checkers are doing. In turn, platforms therefore get a bigger bang for their buck for anything they invest in the space.

Q240 Dr Sullivan: On the point about the most harmful, as we have mentioned in this room, it is the coupling of advertisement alongside harmful messages to the wrong person or to a vulnerable person. How do you define “harmful”? That harmful over a number of weeks or months could then be incredibly harmful, but if you look at the track it may not be harmful to another person. How do we define and pin down what harmful is?

Lyric Jain: There is currently a cyber-security framework in place that the NCSC put out for cyber risks. We could probably follow a similar approach when it comes to information operations, misleading content and other types of harms online where there could be super-concentrated risk around one individual that platforms are required to act on, as well as societal risks that largely go unattended to at the moment where there is a little bit of harm but they are across millions of people.

Q241 Chair: If I understand you correctly, you are setting out a prioritisation of harmful content so that that could be directed towards fact-checkers, and prioritisation would take place in some sort of other algorithm. Who defines what those are? Is there a tension? There are two questions here. Is it technically feasible? You seem to think that it is. I would value Dr Abdul Rahman’s consideration of whether it is technically feasible to have some sort of big AI algorithm looking at everybody’s posts and deciding what is or is not a priority when it comes to harm. Is the question of who decides what is harm not also therefore a question of free speech? Should the Government be deciding what is harm there?

Lyric Jain: That is a really challenging question. We work with a number of allied Governments around the world, and the UK Government tend to be the least prescriptive, and for good reason—exactly because of that tension. There is probably room to be slightly more prescriptive than we are right now in terms of what that prioritisation should look like.

The priority is not necessarily between national security versus public safety. The prioritisation could be based on weighing what potential harms are—you could weigh some harms equally—the likelihood of those harms occurring, and the capability of any threat actor in being able to follow on and conduct offline harm, as a result. I echo Dr Rahman’s points around the ABC framework. It is very rarely just about the content; it is about how atoms and molecules of discourse, as we call them, across content, accounts and activity may individually impact people or in aggregate shape narratives, shape communities and co-ordinate behaviour online. To the point of efficacy on these capabilities, this is something that we have been championing with platforms as well. Last June, there was an International Fact-Checking Network conference. We built something to do exactly that. There are about 30 accredited organisations that have been using that platform since then. It has



HOUSE OF COMMONS

certainly had a bit of a headwind in the last couple of months, given some announcements from platforms.

Q242 **Chair:** Thank you very much, Lyric Jain. Dr Abdul Rahman, can you comment on the feasibility of technically identifying high-risk content?

Dr Rahman: It should be a hybrid approach, and we need to think about what kind of harm. If we are talking about network harassment, in the past when you look at brigading or posts, there were threshold levels that I was told Meta was looking at for Facebook. If a post has more than 30 comments, it goes up to a human moderator to look at. This was around 2019. I think that has changed because of recent developments. It depends on what harm it is and how we define the harm. It should be a societal approach. It cannot be led by big tech. It needs to be in a room like this where the Government, civil society and tech themselves sit together and do it. It cannot be a tech-led endeavour in that way.

Chair: Thank you very much.

Q243 **Jon Pearce:** I want to follow up an earlier question that Emily addressed to Lyric but that I want to address to you, Dr Abdul Rahman. We have talked about the responsibility of the platforms. We talked about the responsibility of the advertisers in the previous session. I am interested in the point about the responsibility of the users. Can a bot be entitled to free speech? Why are the social media companies not taking more action to close down bots at a faster rate? Why are they allowing them to operate? Is the reason for a lot of the harmful content and its amplification that, to take Dr Sullivan's point, in effect, social media has become the wild west? There is no personal responsibility. Is there more we should be doing as a Government in the UK in that space?

Dr Rahman: Yes, completely. I am not an expert when it comes to inauthentic behaviour. This is regarding bots. My research was and still is looking at authentic actors—the Tommy Robinsons of this world and the Elon Musks of this world. Those actors are not being regulated right now. That is a more pertinent question with regard to the Southport riots than bots. For bots, there are methods around it. It can be gamed. We can measure whether this thing is coming online at this time, and then you can find out whether it is a bot and things like that.

The more complicated question is about authentic actors and what it means. It is not just simply that he posts and then this is happening. That is one part of the quantitative method. It is also the qualitative aspect of it—the joy and effervescence of being part of the swarm. People take part in it. There is a thing about the false communications aspect of the OSA that I find very intriguing. How do you also attribute blame when it comes to a swarm or somebody who takes part in a swarm? I say that we cannot do that. We can maybe attribute blame to the amplifier because he or she knowingly does it. This is the problem with community notes. There was a study—I can send it on to you guys—by the Queensland University of Technology that looked at community notes in



the X structure in totality. It is not enough. If you look at disinformation, it is always inflected with humour, and that is where community notes on X fail, because it becomes a reductive binary approach of yes or no, when humour can be weaponised, and we all know that. There is always the possible deniability that amplifiers exercise, so you cannot do intent. That is one part of it. A lot of focus needs to be on the target. There are shockingly very few studies where you focus on the target's perspective. Mine was one of them where I did a perspective to look at what happened to me when I resigned and got all that harassment. Most of the time, we are looking at harassers, so we look at the target's perspective and what can be done and what can be improved. We also help to hold companies accountable for what harms they are doing on their platforms.

Q244 Chair: Thank you very much. Thank you very much, Dr Abdul Rahman, for sharing your experience. We are running out of time. I have one particular question that I want to ask before we move on to the Online Safety Act with Emily. The question I want to ask you is regarding specifically the Meta oversight board, which did not agree to attend today's session when requested. That was obviously a disappointment to us. Can you briefly say what your view is of the Meta oversight board and other independent oversight boards? Can they truly be independent?

Dr Rahman: I am not an expert on the oversight board, but I can tell you more about how we did it at the Twitter trust and safety council. It is quite laughable, but after six years this is all I got. There is no compensation. It is important to have complete independence. This is what we were trying to do under Jack Dorsey. Under Musk, it was a completely different kettle of fish. That is what is important. Having said that, again, like I said before, Elon Musk sets the playbook for how things are happening. Even if we say it is important to have an independent advisory board, companies can just see it as an add-on. The ethos still remains to move fast and break things. If we are really talking about safety by design, they need to move slowly and be deliberate. They need to include people who have gone through online harassment and involve them at the very start when they create a product, not a by-the-way process. I don't think that will happen any time soon. Having a mandatory requirement to have a board, or whatever it is, will be an add-on that does not help the process.

Chair: Thank you very much. Emily, we come to the Online Safety Act and whether that is the solution to some of the challenges that we have heard.

Q245 Emily Darlington: Yes. I have a couple of questions in this area. The first question is a culmination of where we have got to. This is the Science, Innovation and Technology Committee. We firmly believe in science. Is there still room for scientific fact online, particularly when it is contrasted with ideas of free speech? The statements Paul read out earlier clearly have no basis in scientific fact. I appreciate that we are focusing on content here, but where does scientific fact sit within content



moderation? What is the role for enforcement of that?

Dr Rahman: It is a very difficult question. I come from the US. I am based in Washington DC. I think you know the conflagration that is happening there. We are talking about science. I come from the field of public health and development. If you put in a grant right now to the National Institutes of Health and you mention women, you will be disqualified. There are many words that have now been banned: social justice, racial justice, equity, diversity, LGBTQIA, women, indigenous community and native American. There is a complete backlash to science. Sometimes when we talk about labelling and fact checking, it is almost like we are sitting in a room and we are on our own. There is a complete backlash against intellectual ideas. There has been a complete erosion of trust in institutions, so, for me, it is also a question of how we counter that in a way that is effective, because they are doing it really well.

What is happening in politics in the US—there is also collusion with tech—is that it is really performative. A lot of what is happening in political fora is to play to a particular crowd and audience. You see that they use very simple language. We need to do the same if we are progressive and liberal. How do we do that effectively? I don't know what the answer is. Maybe it is storytelling—maybe not so much science as science per se. In covid, in the field of global health where I come from, it is bridge builders. It is not so much your doctor or a GP telling a patient how to take a vaccine. It is a priest or a family friend. Those are the bridge builders who we found were effective during covid in black communities. Whether science plays a role—

Q246 **Emily Darlington:** For this Committee, it is something that is particularly prescient. We are watching what is happening in the US very carefully, particularly misinformation around vaccines, prenatal care or many other medical issues with very serious real-life consequences to young people, women and everyone. It leads to an important question for this Committee and one that has been a bit of an elephant in the room throughout our inquiry. Seeing very active people, including Elon Musk, attack us for being the most regressive on free speech and JD Vance saying that free speech was the biggest threat to Europe at the moment, can the UK really address any of these social media issues and social media platforms without US backing? Can we act alone in this place and be effective?

Dr Rahman: You can build on what has been done in the DSA and add on by working on the thing that I called the vectors—the actors and the behaviours. Meta has been doing it up to this point in time. It is not a new thing for some of the actors that we are talking about. They are familiar with the discussion. When you talk about actors and behaviours, it is protected by the first amendment and compliant with section 230, because that is about content. Once you talk about actors and sequencing and patterns, it is a very different thing. You can make the companies figure out the thresholds. If they do not, make sure that there



is mandatory data access so that we researchers can come in and do that.

Lyric Jain: To build on that point, we do quite a bit of work in the US at the moment mainly focused on foreign interference. What we can do in the UK is probably quite limited without US backing of whatever we are proposing, but there is definitely an opportunity to find common ground, and I suggest that that common ground is around the point of foreign interference and around moving the discussion away from content and focusing on behaviours. First off, foreign interference is one of the only places in the US DoD budget where the US is underspending when it comes to information operations. The US spends \$1 billion a year on countering information operations. China spends about \$8 billion to \$10 billion on information operations every year. Russia spends about half that amount.

Chair: Very interesting.

Lyric Jain: Unless something unusual happens, which it may, given the current politics of the time, that investment in countering foreign interference and information operations as a whole and that impact on national security will likely increase. That is a strong role for the UK to partner with the US on.

On the second point, around behaviours, Meta really framed the language around co-ordinated inauthentic behaviour. It is their definition. It is their word, their term. We should probably step back from co-ordinated inauthentic behaviour to what we are positioning as co-ordinated adversarial behaviour. That includes bot accounts and human individuals who are co-ordinated in a way that is malicious either to a country as a whole or towards certain individuals and certain communities. Looking at behaviours can be quite technical. It can be quite clinical. We do not need to get into debates about freedom of expression. It is content-agnostic. Someone can be promoting DEI. Someone could be opposing it. Looking at the behaviours and not necessarily the subject matter gives us an agnostic and neutral way of intervening in the information environment. Those two are potentially areas of common ground that we can build on, however we work with platforms.

Q247 **Emily Darlington:** A final question from me. Given what you place as important in terms of foreign interference—we all would place that as a significant security threat—should we consider the current JD Vance/Elon Musk campaign against the UK, particularly against the Government and the Prime Minister, and this push about free speech and the misrepresentation of our free speech laws as foreign interference?

Dr Rahman: I don't know the specifics of it. Obviously, I am aware of it. There are some grounds; for example, you can show there is harm. If I go back to the riots, Elon Musk said that civil war is inevitable. You can go back to the data and the access, and we can check to see when he posted that and what happened afterwards. It is not causal, but you can



show with time stamps what happened. With this thing, it is a similar approach where they can show that logical sequencing. For me, it is more from that perspective. Quite honestly, it is your own jurisdiction. You should have every right to legislate it the way you want because you are protecting your country and your own sovereignty, and this is about that space.

Q248 **Emily Darlington:** Lyric, do you want to say anything about that?

Lyric Jain: Not too much to comment on JD or Elon specifically. There is probably a philosophical question of whether the dream of an open internet and a singular internet is now dead. Are we now moving towards slightly differently walled versions of the internet in every country? It would be a shame if we completely gave up on the dream of an open internet that the whole world can communicate through. There has been great power in that over the last couple of decades, and great opportunity lies ahead through AI and other things to keep an internet open. Each country has to make a determination on what its sensitivities are, and it is for that country, its Government and its citizens to come to that conclusion. My view would be that the bigger sensitivity at the moment is in areas that promote violence, or can create offline harms that incite violence, and areas of foreign interference. Those are areas of low-hanging fruit that we could probably work on with any Government or any Administration that is part of the G7 and G20 to try to continue to keep an open vision of an open internet. I would rather not comment on the politics of JD and Elon.

Chair: Thank you very much. Thank you very much for that question, Emily. It is worth emphasising that we in the Committee support the open ability to communicate across the internet. We very much set out in our terms of reference support for freedom of speech. We also support the UK's sovereignty as a sovereign nation. The laws that apply offline apply online. Part of the question is whether some of those laws and regulations need to be adapted for the unique challenges that the online world may present.

You have been very helpful in your contribution. We did not spend very much time on the Online Safety Act. If you have specific suggestions as to how the Online Safety Act or the powers within it could be extended, enforced, adopted or interpreted, please could you write to us? I finish by saying a huge thank you on behalf of the Committee. I am actually more hopeful now, which is not particularly what I expected. Both of you in different ways have suggested areas where the Committee could focus recommendations that could make a positive change to the impact of social media and misinformation in this country, so thank you very much for that.