HOUSE OF COMMONS

Petitions Committee

Oral evidence: Tackling Online Abuse, HC 766

Tuesday 16 November 2021

Ordered by the House of Commons to be published on 16 November 2021.

Watch the meeting

Members present: Catherine McKinnell (Chair); Tonia Antoniazzi; Martyn Day; Christina Rees.

Questions 41 - 76

Witnesses

I: Seyi Akiwowo, Founder and CEO, Glitch; Andy Burrows, Head of Child Safety Online Policy, NSPCC; Stephen Kinsella OBE, Founder, Clean up the Internet.

II: Ellen Judson, Senior Researcher, Demos; William Perrin OBE, Trustee, Carnegie Trust UK; Dr Bertie Vidgen, Research Fellow, The Alan Turing Institute.

Written evidence from witnesses:

– Andy Burrows, Head of Child Safety Online Policy, NSPCC

– William Perrin OBE, Trustee, Carnegie Trust UK

Examination of witnesses

Witnesses: Seyi Akiwowo, Andy Burrow and Stephen Kinsella OBE.

Q41 **Chair:** Thank you so much for coming here today to talk to us about tackling online abuse. I am sorry there not are more Members here today, but several members of the Committee are sitting in Public Bill Committees that are meeting at the same time as us.

Our Committee has received several very popular petitions on this topic in recent years. It is clearly a matter of concern to the public. We originally opened this inquiry last summer, but today's hearing is our second session on the issue since the Government published their draft Online Safety Bill earlier this year. At the previous session, we heard from groups representing communities that are particularly affected by online abuse. We also discussed measures that can be taken to tackle online abuse and how they would impact on free speech, so we are really looking forward to hearing from our witnesses today.

It is possible that there may be a vote during today's meeting. If that happens, we will suspend the hearing and Members will have to go and vote. Do not be alarmed if the bell starts going and I suspend the hearing—that will be the reason. Before we launch into our many questions that we want to put to you today, will the witnesses introduce themselves, please?

Andy Burrows: I am Andy Burrows. I am the head of Child Safety Online Policy for NSPCC.

Seyi Akiwowo: I am Seyi Akiwowo. I am the founder and CEO of Glitch, a charity ending online abuse for women and girls.

Stephen Kinsella: I am Stephen Kinsella. I am a founder of Clean up the Internet, which campaigns to raise the level of discourse online.

Q42 **Chair:** Thank you very much, and thank you again for being here today. Just to say, we do have loads of questions we want to put to you and finite time, so if you can do the best to build on one another's responses where somebody has already covered something, that will maximise the time we have you for today.

There is a question I want to put to all of you, which was why I put that caveat in first. Obviously you represent quite a diverse range of groups and interests regarding this issue. Can you explain briefly what you see as the consequences of the current lack of regulation in relation to online harms? Seyi, I will bring you in first.

***Seyi Akiwowo*:** Thank you so much for inviting me. I also want to just publicly acknowledge and thank Katie Price and Bobby Norris for their petitions around online abuse, and particularly intersectionality of disability and homophobia. The current landscape of the problem that we are seeing when it comes to online abuse is that women are disproportionately impacted by online abuse. UN research shows that women are 27 times more likely to be abused online than men, and Amnesty International's research shows that black women are 84% more likely to be abused online. Girlguiding research that came out just a couple of weeks ago showed that 71% of girls aged seven to 21 are receiving online abuse and are now censoring themselves online. Those are the disproportionate numbers for abuse that we are talking about.

What do I mean by abuse? I mean doxxing, the sharing of non-consensual photography, harassment, cross-platform harassment and targeting on niche platforms—platforms that are meant to be about parenting, animals or animal rights, but are being used to dox people's personal information to troll them. This is having a huge impact on democracy and on censoring public campaigners like Katie Price and Bobby Norris. It is also stifling freedom of expression. Often online abuse is used against freedom of expression as a binary debate—at Glitch we do not believe that to be the case—but we also see a consistent failure of tech companies to design their platforms and new products with gender in mind. Having gender neutral policies and gender neutral content moderation does not help to produce gendered outcomes, which we need. Women do not get the same favourable outcomes needed that men receive.

We are now seeing this with the Online Safety Bill and what is being proposed. I am sure you are going to ask questions about this later, but I just wanted to give you a bit of context on the landscape. The algorithms on these platforms are amplifying and exacerbating gender-based abuse. The Facebook Files in *The Wall Street Journal* had some internal documents from Facebook, and we saw that Facebook is ineffectively tackling hate on their platforms, and is actually amplifying and benefiting from outrage. You have got gender disinformation that is particularly affecting women in politics, such as Luciana Berger. I know Danny Stone gave evidence a few weeks ago that talked about the gendered nature of antisemitism. You have seen this with Diane Abbott and Jess Phillips. You have seen this across the political landscape.

Really and truly, online abuse has been an issue that has not been addressed properly and this Online Safety Bill is creating more gaps. I am really looking forward to discussing some of the ways in which we can beef up the Bill so that it makes sure that women are properly safe online.

Q43 **Chair:** Thank you. I am going to come to both the other witnesses to ask the same question but, since you have led us into this anyway, what would be the one thing that you would like to see the Government do that you think would make a difference to tackling online abuse?

Seyi Akiwowo: I think it is great that we have an Online Safety Bill that is doing a lot for children and tackling terrorism. Children are mentioned 213 times in the Bill and terrorism is mentioned 55 times, but there is no mention of women nor gender at all in the Bill. We cannot leave it to secondary legislation to look at the gendered nature of online abuse. The one thing I would really urge and encourage this Committee to do is to make a recommendation for women and gender-based violence to be mentioned in the Bill. Having the Bill as a tool to mean that tech companies have to take down illegal content quicker does not help women who are facing gender disinformation and are being targeted online, because gender abuse is not seen as illegal and is not yet seen as a form of hate speech. We need the idea of legal and harmful to make sure it is having an indicative list of the gendered nature of certain forms of online abuse.

We also need to make sure that money is ring-fenced in the taxation of tech companies to go towards helplines, services and trauma services to support women who are being abused. You have only got certain helplines around revenge porn—which is an important but niche form of online abuse—and that is dwindling—it is having to reduce its service. We need more variety and specialism to keep up with the beast that is online abuse.

Chair: Thank you. Who would like to go next on the first question, but also the second, if you are happy to go into that as well? Andy?

Andy Burrows: Yes, absolutely. Thank you for the opportunity to give evidence.

Over many years, we have seen that the scale of abuse and the impacts on children and young people has continued to deteriorate. The common thread is that we have seen, time and again, that social networks really fail to engage with the problem. At best, we have seen a lack of investment, a lack of resource and a lack of consideration into user safety and protecting children and young people, as some of the most vulnerable users of these services. But, at worst, we have seen a business model that is designed to produce outrage.

As Seyi was saying, and as we have seen from Francis Haugen, large platforms have business models that are predicated on the basis of looking to push people from centre to edge topics and promoting outrage. The very dynamics—the nuts and bolts of the systems and processes—are designed to deliver precisely that. The consequence is children and young people being subject to a range of deeply disturbing harms that can be life changing and life lasting in their impact. Examples of some of those are online bullying, the tide of deeply abusive content, and the very disturbing trend of sharing nude and semi-nude images with the active intention to cause harm and distress, which is an inherently gendered issue. We are seeing huge numbers of children saying that they are seeing inappropriate content and, at worst, large-scale abusive campaigns where platforms are sticking their fingers in their ears and failing to have any basic systemic processes to identify and then respond to harm—including as harm travels across platforms.

We know that this is a situation that has deteriorated during lockdown.

One in three children told the ONS that they have seen bullying in the last year. Through our ChildLine service earlier this week, we released new figures that showed that we had seen a 25% increase in counselling sessions relating to online bullying during the pandemic. This is something that is disproportionately affecting vulnerable and minority groups. In particular, LGBTQ+ children and young people have seen shockingly high levels of abuse during lockdown. A recent survey from Stonewall found that 40% of LGBTQ+ children had been directly subject to abuse, and there is evidence suggesting that that is around 2.5 times what straight children and young people are witnessing.

As for the point about the Bill—

Chair: What one thing would you do?

Andy Burrows: One of the real concerns that we have is about the scope of the Bill as this relates to children and young people. There is a child use test that has been built into clause 26 of the draft Bill, and this imposes a higher threshold for services to be considered likely to be accessed by a child than comparative legislation, such as that covering the ICO children's code. So a platform will only be subject to the child safety duties and therefore required to take action to prevent exposure to harmful and age-inappropriate content to children if a platform either has a significant number of child users—or is likely to appeal to a significant number of children—and there is a real risk that because of that scope being higher we will not see harm being tackled. We will see harm simply being displaced.

Q44 **Chair:** Stephen, did you want to answer?

Stephen Kinsella: Yes, very much. Building on what Seyi and Andrew had to say, they have obviously talked about the particular abuse that is received by women and children. Opinium did some polling for Compassion in Politics and that showed that a quarter of all social media users right across the across the pitch have received abuse. As Andy said, this has got worse during lockdown. Some research came out yesterday from an organisation called Ditch the Label, which said that since 2019, online abuse has gone up by about 28%. It is still working on this, but it also remarked that this is not just an online problem, because we can see a correlation between, for instance, racist abuse online and hate incidents out in the street. So I am afraid that this is something that bleeds across from the online experience into the real world, if we want to call it that.

In terms of what could be done in relation to the Bill, like everyone here, we welcome the Bill. I think we also recognise that this is probably a once in a generation opportunity to get this right, so we do need to get it right. Melanie Dawes, for instance, in evidence to the Joint Committee, remarked on what she saw as a clear loophole in the legislation at the moment. That is, effectively, that we allow the platforms to conduct their risk assessments and then they are measured against how they perform in relation to their own risk assessments. Generally, we would like to see a serious beefing up of the powers to give Ofcom more power to direct what the platform should cover in their risk assessments, and then compel it to address that.

Specifically, my campaign that we are focusing on at the moment is anonymity. I have some stats I can give you on problems caused by anonymous abuse, but we have a proposal that anonymity should have to be addressed as a specific risk factor. We are not talking about banning anonymous accounts at all, but that it should be identified as a risk factor and something that very much contributes to the volume of abuse. Platforms should have to specifically address that and come forward with plans of how they are going to mitigate it and be held responsible if they do not.

Q45 **Chair:** You would say the challenge is not anonymity in itself; it is anonymous abuse.

Stephen Kinsella: It is anonymous abuse and it is not just abuse; it is also disinformation. The problem when it is abuse is that, because you do not know who is abusing you, it is more difficult for you to assess the level of risk—the level of threat—that you face. When it is disinformation, the problem is more that someone online says, "Well, I am a nurse, and I can tell you that people are suffering reactions to the vaccine and the Government is concealing this," but you have no way of knowing whether that person is a nurse or not. So when it comes to disinformation, our inability to know who we are dealing with makes it hard for us to assess how much weight we should give to what they are saying.

Our proposal is not that we should ban anonymous accounts; it is actually that we should go back to where we thought we were starting with this Bill which is that we were going to take a design-led approach. We should say "Well, what is it about the design of these platforms that makes anonymous abuse and disinformation so problematic?" And we would say, "Rather than try to prevent people being anonymous"—I do not think we should; there are very many good reasons why people should want to be anonymous online—"why shouldn't all of us have the right to be verified if we want to be verified?" It would then be clear to see. Perhaps Twitter's tick is a good model. It should be clear to us which other accounts are verified, and then we could all be given the right to block interactions with all unverified accounts. Give control back to the individual user to curate their experience. In that way, those of us who felt vulnerable to anonymous abuse could screen it out by category, rather than having to block individual accounts as and when they abuse us.

Q46 **Christina Rees:** Thank you for coming along to give evidence today; it is very much appreciated. This question is to all of you to start off with. Do you think the Government's draft Online Safety Bill is ambitious enough on how it plans to tackle online abuse?

Seyi Akiwowo: I do not mind going first. I think it is ambitious in wanting to regulate tech companies. There are many countries in the world that are not doing that. There are more countries that are not regulating tech companies and having those conversations than are. I think it is ambitious in wanting to do that and it is kind of starting from scratch. A brand new Google document is literally needing to be created here, so I understand the ambition is there. I think the Bill currently being proposed fails in being ambitious to support women and it fails to hold tech companies accountable. As Stephen was saying, on a systemic level, it needs to make sure that the tech companies are not just designing new platforms and new ways to beat regulation. It is not doing enough to be holding tech companies to account systemically.

The Bill currently proposed, in my opinion, is also failing to defend democracy. There is no specific mention of the types of abuse that women, particularly in politics, face or gender disinformation, which we know can make and break a woman's electability. This is already really hard and we are already fighting an issue around representation. So, no, I do not think that the Bill is ambitious enough in supporting 50% of the population.

Let me just add to what Andy was saying around how much lockdown and the pandemic has exacerbated the issue. We did a report called "The Ripple Effect" to investigate what it meant when all of us made our kitchen tables and living rooms the new workplace. We actually saw that online abuse against women increased by 46%. When you then looked at black women and non-binary people, this increased to 50%. If we are going to have different waves of lockdowns or remote working is going to be the future, this Bill is not ambitious enough in really holding employers and other institutions to account around a duty of care.

To ping-pong back to Stephen's point around a risk assessment, there is no clear indicative list around that risk assessment, including taking gender into account. An example is Twitter. I declare an interest, if you like, that I sit on TikTok's and Twitter's trust and safety councils. Twitter was releasing a new product around audios—trying to compete, I guess, with Clubhouse. Within 30 minutes of that product being released, women were being abused. Women were being sent pornographic audios, and deaf and disabled people were being completely isolated from the platform. This is something by design. This is something by systems that was not properly checked through a risk assessment and a duty of care that would take into account all equalities. I do not want this Bill to be trying to keep reforming by being amended, and trying to keep up with new platforms. As Andy and Stephen said, we need this to be about systemic tech accountability.

Christina Rees: Stephen, have you got anything to add to that?

Stephen Kinsella: I think I would echo pretty much everything that Seyi said. It is certainly long enough. It is far too complex. Before I started this, I used to be a commercial lawyer—I was an EU lawyer—and I worked on parliamentary drafting. I worked on the legislation that led to the privatisation of our electricity and gas industries. I am afraid this Bill is tortuous. I know you are hearing from Carnegie later today, but they have come forward with a very good proposal to simplify and restructure it. We should be promoting the duties. We should have a foundational duty and we should promote that right to the top and then you should be able to follow it through and see what Ofcom's powers and duties are. At the moment you have to jump back and forth across this Bill to make sense of it. If I struggle—we always say that ignorance of the law is no defence; I think it is on us to try to draft laws that make sense.

There is a lot of work to be to be done in relation to the Bill. The principles are good. We genuinely welcome this. Seyi mentioned the harms to democracy, for instance. I agree. That is a big gap. There is a big gap in terms of how we approach harm and how we approach risk, and then how we direct the platforms to deal with them. I think a lot more structure could be built into this.

Andy Burrows: I would agree with everything that Seyi and Stephen have just said. To add two points to that, right now the Bill does not require companies to discharge their safety duties, or to risk assess on the basis of cross-platform risks and how abuse can spread with real virality and velocity across platforms. To give a very real world example of that, let me talk about a very vulnerable 16-year-old girl called Lily and her mum, Hannah. Lily has diagnosed ADHD and an autism diagnosis. In the past, she had been sexually assaulted. In January this year, videos were posted of that assault and also falsely claiming that she had committed abuse against a young boy. In the course of days, this material spread like wildfire. Within days there were 600,000 views of a hashtag of her name on TikTok. The content spread across multiple sites—Snapchat and YouTube. There was a petition that 30,000 people signed suggesting that she should be prosecuted. Hannah, the mother, was receiving hundreds of abusive messages an hour. There were vigilantes who were trying to track down of the family.

That speaks to the real systemic failing of platforms to identify harmful content that is spreading at scale. But I think it also speaks to the lack of rapid response arrangements where one platform should identify harm like that to vulnerable groups, and then there should be a systemic mechanism to be able to report that. We have seen platforms get their act together on this when it is something like the Christchurch attack, when there are business or reputational drivers that necessitate it. But when we are talking about vulnerable children and when we are talking about families whose lives can be torn apart by harmful content, we just do not see the impetus to action. We need to see the legislation really build in and bake in cross-platform harms through the risk assessment process.

One other thing that I would flag is the absence of user advocacy arrangements in this legislation. As the draft Bill stands, children and other groups at risk of abuse will receive less systemic advocacy protections than, say, passengers on public transport or customers of post offices. With the polluter pays principle that applies across other regulated settlements, we should see levy funded user advocacy arrangements because we need to make sure that this is a fair fight. The regulated firms will significantly scale up their policy and their legal teams to try and influence the regulator and its worldview, to invest in research and to try and skew the evidence base. This needs to be a fair fight and we need user advocacy to be able to speak on behalf of users who are suffering abuse.

Seyi Akiwowo: May I just add to that? I think Andy has made an important point because this is something that was lost in the very first iteration of the Bill—being able to do class action. There was a consultation on if we wanted class action to be a lever as part of the regulatory body—it was not Ofcom at the time—and that has been massively lost. As Andy said, we are already trying to compete with tech companies that are multibillion pound companies. You have the founder of Facebook trying to buy Hawaii, and we are all here struggling to just buy office spaces for our small organisations. We cannot compete.

There is a digital services tax that is meant to generate £400 million a year; 10% of that could be easily ring-fenced towards efforts to end online abuse. That could go towards education or law enforcement. I know you have been hearing a lot around how law enforcement is stretched. A lot of that could go to civil society groups that are providing helplines and user redress. We do not have that. We have an ability to deal with food standards and alcohol, and we can complain at the local supermarket if something has gone wrong. We cannot do that with Facebook; we cannot do that with Twitter.

Facebook has multimillion users on its platform which, when combined, is twice the size of the population of China and three times the population of Europe. Yet you cannot call an emergency service line. We have seen, again from the Facebook Files, how difficult Facebook has made reporting abuse on their platforms over the last year. It cannot be that after Donald Trump has not won his election that it is now putting in the measures it could have put in to stop hate speech from spreading on its platform. This is not good enough. For us mere mortals who are trying to understand the amendment and legal text, and trying to support vulnerable people, where do we go?

If I can talk about my personal experience, I tend to call myself a recovering politician. Back in 2014, I stood for local government. I was one of the youngest black women to ever be elected in local government and a speech that I made at the European Parliament went viral, and I thought, “This is amazing. This is an opportunity to encourage more black women to get into politics.” And then, one day, somebody posted it on a neo-Nazi forum and I was sent death threats and rape threats. I had to do a complete audit of my platform to check that my address was not public. This was not even a year after Jo Cox had been murdered. It was not that I was terrified for myself; I was upset for my mother, who was mortified that her child was being abused. I had to go on TV to get Twitter to respond to me.

It should not take these personal stories. It should not take people who are trying to make the world a better place, who are trying to campaign for society, online and offline, and who put their head above the parapet, for tech companies to take action. So I completely agree that we need redress for users to be able to hold tech companies to account.

Q47 **Christina Rees:** Thank you, Seyi. It is a real shame that we female politicians seem to get attacked. I am not going to share my personal story, but thank you for sharing yours—much appreciated.

Andy, can I take you back to your written evidence a minute? You highlighted platforms' use of algorithms that promote potentially harmful content to others, including young people, as a priority issue for future regulation. Are you satisfied that the framework set out in the Bill will allow for appropriate action to be taken on this particular issue?

Andy Burrows: It is an important question because the use of algorithms—how content is recommended and amplified to children and young people—is a crucial part of then seeing what appears on their timeline. I think what is important is that Ofcom has the investigatory powers and resources that it needs to be able to lift up the bonnet on these companies and to understand how the algorithms are working, how they promote content and what steps are being taken to ensure that harmful content is then no longer being recommended to children. Some of what has been most concerning for me about what Francis Haugen has disclosed about Facebook's practices is the extent to which the algorithms have been actively amplifying harmful and hateful contents, and where we have seen business decisions being taken not to adjust the algorithms in the face of very real world harm. We have seen that, for example, around hate speech. We have seen allegations that that applied in terms of Covid 19 disinformation in the early stage of the pandemic. So we need to see the regulator have the powers and the expertise to get on top of this issue.

One of the other things that I would also like to see the Bill do—this indirectly but, I think importantly, addresses design decisions such as algorithms—is to explore how we can see better personal accountability on people in these companies who are then making the product decisions that eventually inform what a child or young person sees. One of the frustrations for us about this legislation is that it has not learned the lessons of what works well in other regulated markets. In financial services, for example, if you exercise a significant influence function, you are subject to named personal accountability. That is a crucial way of ensuring that if you are someone in a company who is taking a decision about an algorithm, you are thinking that there are personal stakes and there is personal jeopardy, not just the potential for your corporate entity to face a fine at some point after harm has already occurred. I think that is a really crucial way in which we can address the culture through which poor design choices, such as algorithms, then manifest themselves and cause harm to children.

Q48 **Christina Rees:** Stephen, as a lawyer you would probably support that.

Stephen Kinsella: Yes, I definitely would support that. One of the problems I find often in this debate is I sometimes think we underestimate our powers. For instance, I often hear people say, ″Well, these companies are huge. They are global. They are outside our jurisdiction. We cannot do much about them.″ The reality is that the UK is such an enormous digital market. We are not a market that any major platform is going to want to ignore, and we definitely can make laws that will have teeth that will bite here. The companies will not want to have a different business model for the UK. We are actually in the vanguard of potential serious regulation in this field. We are ahead of pretty much everybody else. So what we do here, I think, could very much set the benchmark for how other jurisdictions—including the EU with their DSA—approach it. So, yes, I absolutely agree with what Andy was saying.

Q49 **Chair:** Obviously the focus for us from the petitions that you very kindly paid tribute to, Seyi, that were brought by Bobby Norris and Katie Price, were focused on the anonymity of users of internet platforms. But what you are talking about—Andy in particular—is removing some of the anonymity of the platform creators and lifting that veil as well, which is an interesting take on the anonymity debate.

Seyi, I was also just going to say thank you very much for your personal testimony because I think it is the testimony of the lead petitioners, Bobby Norris and Katie Price, that have really touched people's hearts and minds in wanting to try and do something about this, which is why the petitioners have signed the petitions in their numbers. I think it is personal testimonies that really do help to give volume to those voices, so thank you for that.

Stephen Kinsella: But on the numbers, of course, we have to acknowledge that without social media, these petitions would not have had the success they did. So it is not all bad.

Seyi Akiwowo: Chair, may I just say something about the anonymity of websites. I do not want to name this website, but *WIRED* magazine has done a series of investigations—there will be another one, if not later this week then early next week. They expose a particular platform where you can upload an image: we could take a photo of us today and upload it, and it would nudify us all. We have got no way of knowing who created this platform. We have got no way of holding them accountable. So I definitely think there is a point there around the anonymity conversation being broader than just the individuals to also the platforms creating it. Leading on from Stephen’s point, how do we, as a market, incentivise the type of companies that we want, with safety by design?

At the moment we have this race to the bottom: who can create innovation tech platforms really fast, breaking things on the way? There is this race to the bottom of doing the very bare minimum. You have got Clubhouse that has the minimum of safety policies, yet it has generated millions of pounds in the pandemic. It is the same with Houseparty, the same with Twitter and the same with new products on Facebook. We have the minimum being done right now. This Bill is an opportunity for a step change to make sure that we have a huge standard to which we ask platforms to perform, and to change the dial to say, "A race to safety." You have seen that if you make your safety product or your safety ambition the heart of your company, it works. That is why people are looking for safe spaces in work and to tackle harassment, so why would it not apply online? You can see that with platforms like Bumble where they have put women's safety at the heart of what they do. That is why it is so successful.

In terms of the Bill, we can really incentivise a business model to say, "If you put safety there, parents are going to want their children to be on that platform." There are then going to be more TikTok creators, and more opportunities to engage in democracy and see women in Parliament, and therefore inspire the next generation to want to stand. But at the moment we do not get to see any of that; we just see outrage and abuse.

Chair: Martyn, did you want to ask some questions?

Q50 **Martyn Day:** I think the panel have actually largely answered the bulk of what I was about to ask already, which is probably a good sign because we are all obviously on the same wavelength. Where I was going to go—I will just come in with a follow up to that—was the duty in the draft Bill for social media companies to deal with illegal harms and other harms, and was that prescriptive enough and what should there be. But I think you have all pretty much covered that, so what I am going to ask is: what powers should Ofcom have to intervene or impose sanctions to make sure that if the duty of care is not interpreted properly, we get a result? Would you like to start with that one, Stephen?

Stephen Kinsella: Yes. I would obviously rather we do not have to go down the road of sanctions. I would much rather that Ofcom were given the powers to be very explicit up front as to what platforms are required to do and that they then conducted their risk assessments and discussed before they tweak and launch new products, with more of a consultation about what the implications are. Otherwise, we will always be playing catch-up. The point was made earlier about police and prosecutions. I think we know that the police are already overwhelmed. I do not think we want a whole raft of further offences—quite aside from the difficulties of detection and then prosecution.

I focus more on anonymity and that is what my campaign is looking at. But I would commend the work that Carnegie has done on this in trying to come up with a good hierarchy of duties and also to simplify it. I think it had a concern about whether we really need multiple types of harm to be identified, and then for it to apply to some platforms and not others. Could not we simplify it and have an overriding foundational duty?

Seyi Akiwowo: I completely agree. I think if the focus is on sanctions, it is not really achieving systemic change. We really want to be looking at prevention, not cause. We have been sitting here for almost 35 minutes now. Every 30 seconds on Twitter, a woman is abused online. I cannot even do the maths to work out how many women have been abused since we have been talking about this. We cannot wait for sanctions and then we know that we do not properly have redress. We need to make sure that the duty of care is explicit and—as Carnegie will present later—it needs to be simple.

For Glitch, we believe it also needs to be gendered. We need to make sure that duty of care takes into account the gender abuse that women face. We also think there are learnings from the EU Digital Services Act. This December, the EU exec body will be providing gender-specific legislation to add to the Digital Services Act and I think there is some learnings there for us to be including on that duty of care. I think the risk assessment needs to be about actively reducing abuse, automated decision making around video content, content moderation, and making sure that there is a duty of care around how a young woman—vulnerable, self-isolating or about to embark on her political career—feels safe online and that the algorithms are either not turning on her, or exacerbating already the risks women face by being online.

I think the duty of care to empower Ofcom also needs to have clear instructions and guidance to properly layer of framework in which Ofcom can regulate. My worry is if it is not properly stipulated and it is not simplified, Ofcom could always be in danger of committing ultra vires, because it is not really clear what it should and should not be doing. So I think the duty of care is a really important framework that Carnegie has set out—particular thanks to William Perrin and Lorna Woods on that. I would say to this Committee, in honour of Katie Price, Bobby Norris and also, I am sure, your constituents who are telling you how terrible it is to be online right now, that the duty of care needs to take into account women, too.

Andy Burrows: Just to build on that point of a foundational duty, I think the importance is that this creates a very clear, overarching requirement on companies to identify and then mitigate reasonably foreseeable harm. The danger in such a dense and complex Bill is that we increasingly start to head the other way and this becomes a kind of prescriptive checklist. We have a whole bundle of codes and guidance that everyone, frankly, is struggling to work through. The benefit of that clear, overarching duty—much as we see in health and safety legislation, where there is a clear, overarching objective, and then it is for employers to determine context-specific ways of complying with the legislation—is that it really focuses the minds, and it will force and require the companies to roll up their sleeves and do the hard work to consider the risks on their site and what constitutes an effective response.

On implementing that, for me, it really comes back to the investigatory powers, the information disclosure powers, and then enforcement powers. As Seyi says, if enforcement powers are being used, that means harm has taken place and we are seeing the systems and processes not work as they should. But the deterrent value here is really considerable, so we welcome the fact that there are strong financial sanctions, but we also have to recognise that we are dealing with some of the largest companies in the world. If you consider the cash in hand that a company like Facebook has, even a 10% fine or a proportion thereof is something that they can game and delay for years, and subject to legal challenge. So I think we have to ask the question: how effective is the range of sanction measures that are being proposed in terms of actually hard wiring the idea of a duty of care and viewing this as a kind of compliance piece at every stage of the business?

I think that is where we need to see more. Again, for the NSPCC, that brings us back to the importance of named responsibility. How great would it be if, through the Online Safety Bill, each of the companies had to have a named person who was responsible for harm against children and harm against women? There would then be direct personal accountability for the actions that then translate into harm that is being caused to children, women and families up and down this country every single day.

Q51 **Tonia Antoniazzi:** Again, a lot of what I am going to ask has been covered. This question is really for Seyi. You have called for the disproportionate abuse faced by women and girls to be specifically recognised in the Online Safety Bill. It is quite worrying to see the word "woman" not there—it is appalling. But what is this going to look like in practice? You have spoken about a duty of care, and Andy was talking about compliance. Is there anything else? What would it look like? What will it be?

Seyi Akiwowo: What it would look like would be involving civil society from the beginning in some of the decision making around the platform. There are trust and safety councils, but civil society groups—some people might disagree with this—are not properly remunerated for being on those councils, so how can they compete with the policy minds? How can they invest in their organisations to sense-check that what they are being fed is actually true and not just the PR spin? Civil societies are not properly armoured, if you like, in the battles with tech companies when the door is slightly ajar to kind of see what is going on. I think what it would look like is civil society being a part of tech companies' trust and safety councils and oversight boards. Again, we have seen with Facebook Files that oversight boards can be lied to, but that is at least what it would look like.

I think it would also look like Ofcom working with civil society groups very closely. I know that Ofcom is hiring, and I think it is definitely building its capacity, but we cannot expect it to be ahead of the curve all the time. There are civil society groups like Glitch, NSPCC and EVAW, which are providing on the ground support to women. Glitch provides online safety training to women who want to be online. We support them with their digital safety, digital security and digital self-care. That has given us a real ear to the ground on new forms of online abuse, and we can therefore be ahead of the curve and able to improve our advocacy. This is something that I would love to start seeing with Ofcom.

Putting my recovering politician hat back on, and going back to Andy's point about a case review, I sadly remember when we would do case reviews when there was homicide, domestic abuse or child safety safeguarding that had gone wrong. There would be a whole council case review to look at what the failings were. We do not have that with tech companies. Every day that somebody is being abused online, our tolerance level increases once again and we now move from looking at abuse to violence. When we start looking at violence, it is too late.

The third thing is tech companies' reporting on online gender-based violence. At the moment, many of them volunteer to do an annual report but, again, a lot of that looks like PR spin, so you have to read behind the stats. I think it would be really interesting to have a clear breakdown of how many accounts had to be taken down in X period of time that were antisemitic, anti-black, homophobic or ableist, or that had an impact on democracy, and then let us assess them on the same indicators in six months. At the moment the indicators and metrics change and, as Andy said, civil society groups are working on a shoestring budget and cannot keep up.

Q52 **Tonia Antoniazzi:** We heard suggestions in our last session about important limits on the functionality of unverified accounts, and that could mean that vulnerable users who are unable or choose not to verify their online identity could lose their ability to contact MPs or other high-profile figures. Can that risk be managed or is it just the inevitable trade off? It really is an issue because we were hearing in the last session that anonymity to some people is key for them to be able to be in contact.

Seyi Akiwowo: I will hand the mic to Stephen on this for the work that he is been doing on anonymity. What I will say is that a lot of women need anonymity and pseudonyms to participate online. With gaming, for example, the Gamergate that took place in 2010 and onwards saw an exodus of women not being able to be themselves online because they were just too good at FIFA and The Sims—you name it, all the games that are out there. So sometimes they are a protection—literally life and death for some people.

I think what we should also be seeing in the risk assessments, to make sure there is a duty of care at the heart of this, is that we are investing in safety tools. At the moment we only have blocking, filtering and muting. I think that that could be expanded so much more to give users a whole range of tools to have proper agency on the platforms so that they can opt in to say, "Okay, I am about to speak in front of the Petitions Committee. This might spark abuse. I'm going to able to turn up my security settings for the next couple of days so that I don't have to see that." We do not see enough being invested in safety tools or reporting mechanisms so that people can opt in and out when they want to when it comes to anonymity.

Stephen Kinsella: I know in one of the previous sessions some of your witnesses talked about the idea of stable accounts, did they not? It was the notion that over time an account would acquire a certain status and then would have more permissions—I think that is what you were alluding to. Our proposal does not go that way; it goes completely the other way. We say that an unverified account has the same status as any account. They do not lose permissions or abilities to do things. All we say is that each one of us should be able to choose for ourselves not to hear, not to interact and not to receive replies from unverified accounts, and that would also have a benefit. It would not just mean that I would not see it. Let us say that Marcus Rashford could communicate to everyone out there who likes to follow him, but if he decides to activate this permission, only those who were verified would be allowed to reply to him and therefore he and also all those who follow him would not see those replies. That would greatly diminish the ability for people to disseminate abuse.

The MP question I have heard come up a few times and it is a good one. I am not sure how many MPs would encourage their constituents to bring matters to them via Twitter. When I email an MP, for instance, I always get an automated reply saying, "First of all, I can only deal with you if you are a constituent, so can you provide something to verify that you have an address in my constituency?"

Q53 **Tonia Antoniazzi:** I think the example is that if there is somebody in danger, the only means that they have is to get hold of you on your Twitter. I think there were certain issues that were spoken about. It is not the norm that I would talk to my constituents via Twitter direct messaging, for example. We would say, "Please go to my email account." However, there are cases where people need to.

Stephen Kinsella: Yes, of course. That would have been you as the recipient of that message, so it would have been your choice to decide that you did not want to hear from unverified accounts. You might say, "Well, I take a view as an MP." I have obviously spoken to a number of MPs about this. Margaret Hodge receives a great deal of abuse, but she says she wants to know about it. She does not even want her staff to screen it from her. Obviously many others use staff or people to filter, but they say, "I wouldn't stop any messages coming through to me, but I would have a screening mechanism for it." That would obviously be a choice on an MP-by-MP basis.

When we come to whistleblowers, again, I am not sure that they would use Twitter very much as a mechanism. If you wanted to communicate with *The Guardian*, for instance, and you want to send material that you think it should follow up on, there will be a secure drop box. I think we always have to be very sensitive, of course, to the risk of unintended consequences. What we are talking about today is, after all, a lot of the unintended consequences of these business models that really ran out of control and where that "move fast and break things" ethos dominated everything else. I acknowledge that.

We have tried to think about what the real downsides would be of simply saying to each one of us, "You have the right if and when you choose." For instance, it could be that you might have been in the headlines for something and you think, "Well, for a couple of weeks, I think I will just dial this down. I will not take all the unsolicited material from people who are completely unverified, and then once that has passed, maybe I'll dial it up a little bit again and see what happens." But it would give you the choice as the potential recipient. The thing with abuse is it has two elements. If the abuser is just shouting into the void and nobody is hearing them, they are probably still technically committing an offence, but they are causing far less harm.

Again, I think our proposal would have a great benefit in reducing the number of harms that the police or other authorities might have to follow up on. I have discussed this with the victims commissioner. I have discussed it off the record with prosecuting authorities as well. They can all see the benefit in terms of their workload and perhaps even in terms of your workload.

Seyi Akiwowo: Just on the point of harm, which is why we have been looking at a public health approach to addressing online abuse, I think there is scope in the education arm of the regulatory powers that Ofcom will have to really be looking into this because, as Stephen said, it is not just harm of the person who is facing the abuse. It is their friends and family, and those who see it—it has this ripple effect. We know from youth violence that a public health approach and deeming it as a public health issue really helped to address that ripple effect of when one person has been stabbed or been abused, and the impact it has on the community and the school. I think a similar approach when it comes to risk assessment is harm prevention and harm reduction, so that not everyone is seeing it.

If we look at the characteristics of a troll—I am fascinated about people who become trolls and very obsessive online—what they really want is to be seen. A particular tactic of abuse is called ratio. For example, Chair Catherine, I might tweet you later and say, "It is been really lovely to engage with you. Thank you so much for the invitation," and then we may have a friendly dialogue and post a selfie. Ratio would be where trolls are maliciously trying to have more negative replies than the likes and the retweets—the positive engagements. That means they want to be seen.

Glitch did a campaign with BT Sport earlier this year on drawing the line in sports, because we were seeing abuse affecting the game of sports. Some 1.8 million were affected, and half the UK population had seen online abuse. If we are really looking at a public health approach and making sure that we are equipping citizens online to be digital citizens being responsible with their platforms, we are then reducing the amount of abuse people are seeing, which we know has a knock-on effect. One in seven people who were surveyed expected people in public life to have to face abuse. That means there is a massive disconnection around what harm is, because we have been subjecting people to harm overload for so long. So there is a real power to make sure that we are teaching people to use anonymity as a tool for good and for being responsible, as well as what a public health approach to tackling online abuse would look like.

Q54 **Chair:** We are going to come back to the anonymity and explore some of the proposals that you were talking about a little bit, but first I just wanted to come back to you, Andy, about children, because often I think a lot of this debate does understandably revolve around adults and how we manage in this online world that has normalised abuse—Seyi very well painted a picture of that. Are you concerned that children have been a bit overlooked in this discussion, or do you think we are sufficiently addressing some of the specific risks to children and young people?

Andy Burrows: Anonymity is very much a double-edged sword when it comes to children and young people. There are plenty of examples of where being able to speak with an anonymous voice is important for children and young people. On that basis, we would not support an outright ban on anonymity. We would very much endorse the approach of dealing with this as a risk to be mitigated and that is very much through some of the design features that Stephen has been speaking to. Some of those may be bespoke or play a particularly important role in relation to children. For example, are there steps that platforms could take to introduce some friction into the user experience so that children can have more control? For example, if a child is contacted by someone and they choose to accept a friend request, could there nevertheless be a cooling-off period of 24 or 48 hours so that then, if that account is aiming to be abusive, at least there is a kind of cooling-off period.

There are some very novel design features looking to build this friction into the user experience, which it is reasonable to be expecting companies to consider as part of how they discharge their safety duties. For us, this is about how we can mitigate the risks of anonymous abuse. There is a danger, particularly to vulnerable children, of not being able to speak with anonymous accounts. There is a danger of losing the baby with the bathwater.

Q55 **Christina Rees:** Seyi, you have argued that digital citizenship and behaviour change should also be key elements of the response to online abuse. What should this look like in practice in terms of education in schools, but also to reach adults?

Seyi Akiwowo: That is a great question. The concept of digital citizenship is important because, again, it is about systemic social norms and values that we can all have when it comes to the online space rather than designing a curriculum to keep up with new platforms. When we set up Glitch, TikTok did not exist, so can you imagine trying to keep up with different platforms and providing a curriculum on that? Digital citizenship is a concept of rights and responsibilities, and what it looks like to be responsible online. As an individual, it is thinking about your digital security: safe passwords; two-factor authentication; and making sure you are not logging into certain sites with your social media accounts. It is being responsible with your devices, which is also an open window to other people—your friends and your family.

Digital citizenship on a social level is around what we do when we see someone who is facing abuse online—someone who is on the receiving end and being targeted. We know instinctively if we see a fire to call 999, and we know we know instinctively what to do when someone needs medical attention. We get first aid training. There are so many ways in which we get standard training when it comes to instances offline; we have not begun to have that conversation online.

On an institutional level, we would argue along the lines of what Stephen and Andy were saying: products and tools that nudge and encourage digital citizenship behaviour, that encourage communities and that also support people who are running Facebook groups in the hundreds and thousands. There was probably a massive increase in people creating Facebook groups, WhatsApp groups and Signal groups during the lockdown, because there are a way of keeping community, which is amazing. All it takes is one person being rude, disrespectful or abusive, and then the whole community can crumble. How are platforms supporting group moderators with how they can set out principles of how to run a group, how to run a campaign, and making sure people are signing up to those kind of community guidelines within a group so everyone is on the same clear page?

We have seen what it looks like when we have public health education or public education campaign that focuses on everyone playing their part. We saw that with drinking and driving. In the 2000s, there was a heavy focus on everyone plays their part in keeping the roads safe. We had 20’s Plenty and do not drink and drive. Peggy Mitchell would be taking people's keys when they were drinking too much in the Queen Vic. We had a real investment on billboards and TV around our all needing to keep our roads safe. We saw investments in seatbelts and around the importance of car seats for children. We do not have any of those safety conversations or those bystander conversations when it comes to being online. That is what digital citizenship is.

Particularly for women, we have a programme on digital self-care because it is really hard to set boundaries online, particularly for women in politics. It is often said, “Well, it is just part of being a woman in politics," or, "It is just part of being in politics—get out of the kitchen if you can't stand the heat." Those are very much victim-blaming narratives. Digital citizenship is about countering that, and not passing the blame to somebody else and making the victim have to deal with it. Digital self-care is about, "How can I have boundaries online? How can I flourish and have agency?" I really hope that in three or four years' time, when this Bill has passed and is hopefully successful, we can start having a conversation around what joy looks like online and what pleasure looks like online. We are so focused on trying to reduce violence, harassment and death that we have forgotten that the online space is meant to be for innovation. That is also what digital citizenship is about—using these platforms responsibly.

We provide training and we saw that 100% of participants that went to our digital citizenship offering gained new skills. Some 89% felt able to intervene confidently and safely, and 85% felt more confident to use the safety tools. When you sign up to Twitter, Facebook and Instagram, no one gives you a tutorial of how to use the platforms, so no one is giving a tutorial on how to use it responsibly. That is what digital citizenship education is about.

Andy Burrows: Could I just briefly come in there? One of the things that is very characteristic of the debate around the Bill is very keenly held and legitimate concerns about how the Bill stifles freedom of expression. But freedom of expression is also the right to participate. It is the right to be present in online spaces. What we are seeing is children and young people, women and other groups who are faced with the decision of either having to absorb absolutely unacceptable levels of abuse and negative experiences online, or withdrawing. We see that when we are thinking about pile-ons and harassment—is there a willingness to engage in conversation? Because then there is the fear of the trolls, the ratios—all of those things coming out. I think we need to approach these issues by recognising that there is a real prize of actually safeguarding and securing freedom of expression, and that should be an objective as we deliver this legislation.

Q56 **Chair:** We will go to Tonia, who wants to ask some questions, but I said, Stephen, that we were going to come back to you on anonymity and how we get that balance. I think there is a bit of a debate about whether we value anonymity for the freedom it gives people online, and how you get that balance between creating safe spaces where people can participate and your proposal about adding in the verification element. Really, the real crux of the debate is how you get that balance right. I guess some would argue that your proposal is infringing on your freedom of speech and your right to anonymity online, and your proposal is a bit like how we all try to stop unsolicited mail coming through our mailboxes, but we still end up with a massive pile every time we go to collect our mail. We have not achieved it in the physical world; perhaps it is possible to achieve it in the online world, I think, is what you are proposing.

Stephen Kinsella: That is a good comparison because it would be exactly that. We would not be saying we will ban the companies from sending the unsolicited mail. We would not be saying they cannot produce it and print it and try to send it to us. But we are all able to put, as you say, on our letterbox on our front door, something saying, "No circulars; no unsolicited mail." Nowadays you can also instruct the Post Office so they have to take note of that and not deliver it to you as well. Now, maybe that is not always as effective as it should be but, if there is a free speech right at all in the right to send unsolicited mail, it is not affected at all. We are just saying each user should be given the ability to decline it.

We are very conscious that one of the issues that people get concerned about is, "Well, what do you mean by verification? How would that work and what are the risks of that?" There is obviously a great reluctance to give yet more data to the platforms. There is a lack of trust about what they would use it for—whether they would use it to target us and monetise us. But there are some very good third-party solutions out there. For instance, there is a company you may have dealt with called Yoti, which provides verification for the Post Office. There is a company called One ID, which uses the data that we already have with our banks, so it creates a bridge between the bank and the platform. You have already gone through the exercise once of verifying yourself to your bank. It then generates a code and you can then use that to open your social media account. It has the great benefit that it does not give them the data. That is, of course, one of the reasons why it will not appeal to them. I know that some of these companies have tried to get the social media companies to work with them and they do not.

I think there are three reasons. One is they get denied that data. The second is what happens if we move to verification becoming more common. The Opinium polling that I mentioned for Compassion in Politics showed that more than 80% of users polled said they would like to be able to be verified if they could, and a similar percentage would then exercise their right. I think it was 73% who would then exercise their right not to have interactions or unsolicited communications from unverified accounts. So as you move to verification, it will become apparent how many of the users of these platforms are actually illusory, how many followers are fake followers and how many accounts are fake accounts. That would obviously have an impact on their advertising revenue, I would say, in the short term.

So, for verification, I think there are good third party solutions out there that one can trust. Let me just give an anecdote. We are asked by the platforms to trust them. One of the problems I have with the Bill is that a lot of it does rely, as we said, on the platforms doing their own risk assessments and there is a great reliance on them doing it properly and scrupulously. Well, you will recall the Euro finals, when our footballers missed those penalties and the amount of racist abuse there was. Twitter came out with a surprising claim: it said that 99% of the accounts that they took down were not anonymous. Now that just surprised us, and it would surprise anyone who has spent more than five minutes on Twitter. I think it over-claimed—I do not know why it went for 99% to be honest—so we challenged them.

We wrote to Twitter and it ignored us. We wrote together with Kick It Out and it ignored us. We got Margaret Hodge to write, and on the second time of asking from her, it finally replied and confirmed what we suspected: its definition of what anonymous means is very unusual. It says that your Twitter account is linked to an email address or a phone. Well, we tested that. We set up an account called mickeymouse@gmail.com and we went into a newsagents and bought a £10 pay-as-you-go phone for cash. Using that email address and that phone number, you can open a Twitter account and it will say, "Of course it is verified because it is linked to an email address or a phone number." So if you get the platforms before you, the only thing I could say is treat their claims with caution. I would certainly ask for verification of the claims that these platforms are making about their attitudes to security.

Q57 **Chair:** We have run out of time, but it is interesting what you are saying because we have taken this inquiry out to schools in some of our constituencies. For me, the main messages that came out from the young people we talked to were that they are crying out for digital citizenship. One suggestion that they made was, "Why cannot we just link all internet accounts to a bank account?" And this came from the young people—they were not things that I had particularly said to them—so it was quite interesting. It was just some of the feedback that we had from a young person's perspective. Do you have any final comments before we go to the second panel?

Seyi Akiwowo: I know you are short on time and I definitely believe you need to be hearing a lot from Carnegie in your second panel, but I would say, to mirror or complement what Stephen was saying, that we do need to be looking at offline education. People are using anonymity and saying these hurtful and abusive things, and then being radicalised to join movements like incels online. We need to be looking at what is driving that offline. We need to be looking at our curriculum and our progressive education offline that is meaning that people are holding these abusive and harmful opinions, and then feeling emboldened to it say online. What are they learning offline? They have been hearing it from politicians in the US, from celebrities and from role models. As much as this is an online issue, online violence is a continuation of offline violence, and so with anonymity and tackling that, we need to be looking at why people have those behaviours. And there are great organisations on decolonising your curriculum that are really tackling some of these values that we should be having for each other that I would invite you to speak to speak with.

Andy Burrows: Can I just make a very quick final point about transparency, because I think this is a crucial part of what effective regulation has to look like? As an example, Facebook's self-transparency reports suggest that between 0.14% and 0.15% of posts contain content around hateful speech and bullying behaviour that would breach their terms and conditions. But, of course, that is based on a general user; it is not based on someone who is subject to targeted harassment. It is not based on a woman in that situation or an LGBTQ+ child in that situation. One of the things we really need to see is transparency—and granular transparency—that speaks to the user experience, because data is only as good as we are able to interrogate it. When we are seeing, at best right now under self-regulation, those global figures, they are actually pretty meaningless because if you are then the child or young person who is on the receiving end of sustained harassment and abuse, those global figures mean nothing. That means that then you are going to bed at night having absorbed online abuse all throughout the day, and that abuse that could be prevented—it is inherently avoidable.

Stephen Kinsella: May I make a very brief final point? We have been working very closely with Siobhan Baillie MP and, as you probably know, she has a ten-minute rule Bill next week—Wednesday 24th—which will put forward our proposed language on anonymity. I think that that will be a good way of testing the breadth of parliamentary support for that approach.

Chair: Thank you so much for your time today. We really appreciate it.

Examination of witnesses

Witnesses: Ellen Judson, William Perrin OBE and Dr Bertie Vidgen.

Q58 **Chair:** Thank you so much for joining us; we are very grateful. Could you first of all just introduce yourselves, please? Ellen, do you want to start?

Ellen Judson: My name is Ellen Judson. I am a senior researcher for the Centre for the Analysis of Social Media at Demos. We are a cross-party think tank and CASM is our dedicated digital research hub that focuses on tech policy.

Dr Vidgen: I have got a couple which I will go through. My name is Bertie Vidgen. I am a research fellow in online harms at The Alan Turing Institute, where I lead the online safety team and work on the online harms observatory. I am also co-founder of a small start-up called Rewire, where we are building socially responsible AI and safe-by-design systems. I also advise the online safety data initiative, which is a DCMS project, to improve access to data for online safety. I am also a special advisor to the Online Safety Bill pre-legislative scrutiny Committee. Some of my collaborations do involve researchers from tech, but everything is open research with open data.

William Perrin: I am William Perrin. I am a trustee at Carnegie UK Trust, where I lead a team of people who have been working on online safety issues since 2017-18. We originally devised the statutory duty of care approach that has been broadly adopted by the Government in the UK and, to some extent, in Europe in the Digital Services Act.

Q59 **Chair:** Thank you all for being here this afternoon. I will direct my first question to you, Bertie. Your research, as you mentioned, looks at how AI can help to detect hate speech online. What advances do you think this technology in particular will have in trying to tackle online abuse, which is clearly what we want to try and get to the heart of in this inquiry?

Dr Vidgen: That is a tough question. I think one of the first things we always encounter is why you use AI and actually, over the last few years, we have seen so many issues with the use of AI, so it seems to not pick up on the type of abuse and hate that we want it to, but it also seems to take down legitimate expressions that we do want to leave up. It also seems very brittle. It has lots of blind spots. It has lots of biases and so many issues with its performance. It does kind of raise the question, "Well, why do you even use it?" And the answer to that is actually very simple: it is just a question of scale. The sheer volume, complexity and variety of online content that we are seeing means that, especially if you want to do proactive monitoring, you just do not have an alternative.

I sometimes say AI is the worst way of moderating platforms, apart from all the others that have been tried. Because the alternative at the moment is to have human moderators sitting through and reviewing all of that content, and then having to make decisions about it. That is very difficult for them. It inflicts a lot of harm. They are also very inconsistent and it does not scale. Let us say you have an unexpected event—perhaps it is a terrorist attack—and then the amount of online abuse massively skyrockets. Well, if you have only got 100 moderators, okay, they can each to do a bit of overtime, but that is not going to scale to deal with the problem. AI can. So there are lots of reasons why we use AI, and I think the question now is really not should we use it, but how do we use it and how do we have the right oversight. There are actually very few people in this space who have thought about this seriously, or have been here more than five minutes, who think that we should have an AI-only system. That is not really being put forward by anyone. I do not know any platforms that have that or are moving towards that. I think it is all about a hybrid flexible approach.

Some of the big challenges that we are going to have to overcome with AI actually relate to freedom of expression and also privacy. How do we build AI systems that do not harvest more data? One of the big challenges right now, for example, is that context is hugely important. Who you are, who you are speaking to, the social setting and the platform setting all make a huge difference. How do I encode that into my AI? Well, do I want to start having an AI that understands the identity of the speaker? Okay, on the one hand that could make sense if you are talking about the use of the N word. That is incredibly different if you are black compared to if you are white—it just has incredibly different resonance. But do I want my AI now to be able to infer or understand your identity to make a better assessment? I think the answer is no. There are some really, really tricky questions that we are going to face if we want AI to be better.

We should always be aware that it is not trivial to build good AI, but the really difficult question is what we want the AI to do. Where do we want to draw the boundaries between, say, hate and not hate? What is permissible and what is not allowed? And how do we draw that in a way that is consistent and clear? Then we can go and build AI that reflects that decision boundary, but if we do not know what the decision boundary is—if we have not made that decision—we are never going to have AI that does what we want it to do.

Q60 **Chair:** What has your research suggested might be the most effective way of enabling platforms? I know you proposed a lot of questions in your answer there. I am wondering to what extent you have actually reached any conclusions on the answers. I guess I am thinking that if the legislation creates this overarching duty of care, AI presumably is going to play some role for tech companies in meeting that duty and obligation. How do they get the balance between the right use of artificial intelligence and human moderation? And how do we make sure that when that balance is not right and therefore they do not get the right outcomes, you cannot just hide behind the excuse of, "It was AI; it wasn't a human error"? How do we get that balance right?

Dr Vidgen: Anyone who uses AI should be responsible for its use. It is no excuse to say, "I used AI and it spat out some magic answer and I am now just going to accept it." That is not okay. One of the key things we can have is user challenges. Users should always be able to challenge any decision that they have been subjected to. The other one is that we want explainable AI. It is a really big problem right now, and actually this is true even if it is a human-led process. If you are on a platform and they kick you off, they often do not tell you why. In some cases you can probably work out why it is, but in lots of cases you will not be able to, especially if a mistake has been made and legitimately you cannot work out what you have done. They have not told you what policy you have contravened. They often have a lot of policies, so it is not easy to make sense of it. So unless we can get explainable content moderation, it is going to be very hard to enable people to make the challenges to when the system makes mistakes, and that is absolutely key.

The other thing is trying to be problem-driven. So, rather than saying, "Right, we want AI because AI is shiny and new and fantastic", it is, "What is the problem that we are trying to solve and it is AI the best way of doing this or is AI just one tool amongst many that we should be using and we should think very carefully about it when we do use it?" We will probably get on to this separately, but I think this really also comes down to safety by design and how we build safe systems, rather than just trying to tack a bit of content moderation on to the end of it that we think will solve all the underlying problems. If we have not built a good system to begin with, it is not going to do it—it is like putting on a plaster. It is just not the right answer.

Q61 **Tonia Antoniazzi:** Ellen, you suggested that action on online harms should focus on encouraging social media companies to change features on their platforms that make harm to users more likely, rather than trying to ban specific harmful content. What are some of the key changes of this kind you would like to see the platforms make to address online abuse?

Ellen Judson: I think there are a lot answers to that question. What we want to be seeing is focusing on what systems platforms have in place that goes much beyond content moderation systems. Certainly content moderation systems should be part of a risk assessment and an assessment by the regulator as to what steps platforms are taking. But content moderation systems, as they currently are, are always kind of post hoc. They always wait for a problem to occur and then it is: can you report it; is it taken down; are you banned? What we would like to be seeing is much more proactive steps being taken up front by platforms to design systems in ways that reduce those risks overall.

Some of that will be related to what Bertie was talking about with AI content curation, for instance. It might recommend to the systems the way that newsfeeds or timelines are created and served to people. I think there are also questions around the kinds of communication that users are able to engage in with each other. There is a lot of concern about throwaway accounts, for instance, where people just make an account to abuse somebody and then delete it before moving on to another account. It is having systems where you cannot just post immediately. If you make an account, you can immediately message anybody you want on the platform. I know some platforms like Reddit have systems where you have to build up a bit more of a community identity. I do not mean identity in the verification sense, but in that sense of people on the platform know who you are, can see what you have been up to, and what other things you participated in online.

I think there is also a bit of an evidence gap around specific design features when you are getting into the really detailed elements of what we want to see. We have obviously seen that from the Facebook Files. What the whistleblowers have been sharing over the years is that platforms are doing these kinds of tests. They are tweaking things like the algorithmic systems. They are making design changes. Twitter increased its character count. Did that have any effect on abuse? I do not know, but I feel like that is the kind of question that we would want the regulator to be able to ask—"I have seen you have made this change." Seyi was talking earlier about the introduction of audio on Twitter. What is the effect of these little tweaks when they are made? Even if there is not an obvious connection to, "Oh, I can see straight away how that would lead to greater risks of harm," I think it is a question we need to be able to ask the platforms and not just rely on their telling us, "Everything will be fine," which is what they tend to say.

Q62 **Tonia Antoniazzi:** Yes, that accountability is needed. We also spoke earlier about the unintended consequences of what they are doing as well. You have spoken there about the systems. How could the duties in the Bill be adjusted to put more emphasis on system design?

Ellen Judson: I think at the moment there are a lot of elements in the Bill which are focusing on systems, but they are sometimes a little bit secondary. On the legal but harmful duty, for instance, it is focusing so much on terms and conditions and the enforcement of platforms' terms of service. I think that is one important system. But again, that is going to be something that is usually going to be reactive after something has occurred: "Are you taking it down? What are you doing about it?" I think there are mentions elsewhere in the Bill that this will include algorithms, design, platform design and business models, but I think more focus on how that could be brought more to the front of the duties and not as a kind of secondary point after the content moderation is really what it is going to be focused on.

I think a lot of the worries that we have heard about the Bill from folks particularly concerned about freedom of expression are coming from this place of, "Well, is this a Bill that is going to enforce content being taken down?" and what is wanted is just content being removed, and that is the change. Then there are a lot of worries about how we balance that with all these different freedoms. But when you are thinking about design processes, obviously freedoms need to be taken into account, but it is less of an immediate threat at that level.

Q63 **Tonia Antoniazzi:** Do you think the Bill strikes a good balance between requiring action on specific types of harmful content and encouraging companies to change the features of their platforms which are going to increase the risk of harm to users?

Ellen Judson: It is difficult at the moment—before the priority harms and everything has come into play—to predict exactly what changes the Bill is going to require platforms to do. It would be good to have more clarity on how far platforms are going to be assessed on their actions about specific harms versus just generally, if there are common systems. For instance, an algorithm that is promoting divisive content may not be linked to a specific harm, but it can be linked to general risks of violence, amplification of violence, amplification of abuse, and so on and so forth. I think the risk of pinning down exactly the harms that we want addressed is that if anything that does not fall exactly in that scope, platforms might be able to say, "Oh well, it is not our problem." I am thinking particularly of things like disinformation, where often there is a fine line between disinformation and abuse online. They can be commonly connected, particularly in attacks against women in public life. Our concern is that if societal harms and those broader misinformation and disinformation elements are explicitly excluded, it gives platforms a way out by saying, "Oh, but that is not in scope. We do not have a duty to do anything about that," when it may be that similar design changes would help both problems.

Q64 **Christina Rees:** William, you have argued that the final Online Safety Bill should include an overarching duty of care on platforms to protect users from reasonably foreseeable harms, such as online abuse. Why do you think this is necessary?

William Perrin: We think that the current Bill is split up into so many small pieces in order to constrain the operation of the regulator—to keep it very tight and very focused, and to prevent regulatory creep—that there are gaps opening up between a variety of different duties within the Bill. As Ellen has said, the scattering throughout the Bill means it is very hard to get a holistic view of where the harms sit. In some senses it is a bit like the Grieg piano concerto sketch: it is many of the right notes, but not necessarily in the right order. This adds greatly to the Bill's complexity, which increases its regulatory burden and also makes it very hard for victims and victims' representatives to reassure themselves that the Bill will do its job. A few days ago, Carnegie published a substantial reworking and reordering of the Bill to structure it such that it opens with a very clear set of objectives. It gives powers to the regulator and then brings in a broader definition of harm and a more comprehensive statutory duty, which we style as a foundation duty. We do not believe it should replace the specific duty on criminal issues nor the specific duty on children, but it would serve as a stronger foundation for them and also underpin issues around harm to adults. I will just explain a little about how we see that working.

We feel that the regime will not work well unless you have an almost complete understanding of the range of risks that are arising, particularly on the major platforms. In the current regime—as set out in the Online Safety Bill which is designed to be very tight, very focused and restrictive—it is not clear that Ofcom, or indeed the companies, are really tasked to do a very comprehensive and systematic risk assessment of all the risks that might reasonably foreseeably arise during the operation of the platform. The language in which many of the risk assessments are bound is a language of proportionality. It is a language of economic regulation rather than of safety regulation, which we see, for instance, in the Health and Safety at Work etc Act 1974 when it talks about reasonably foreseeable risks of harm arising.

So we have pulled all this together into a duty that looks at harms arising broadly through the operation of the platform—not necessarily on specific types of content, but arising through the operation of the platform. So it captures the systems, the processes and the way the platform is run, and that includes, to some extent, of course, the content within that, but it is not focused upon it. We bound a duty with a sense of appropriateness. Nobody wants a totally bland internet that is completely safe and where no one can take any risks, so in some cases it might be appropriate to have different tolerances of risk. We also look at talk about preventing and mitigating harms arising, which is language from the United Nations *Guiding Principles on Business and Human Rights*, which suggests first we have an obligation to prevent harm and, if that has not worked, you have to mitigate harm that arises. We also link it very firmly to risk assessments. There seems to be a gap in the Bill between Ofcom's broad outlining of risk assessments that it expects companies to do, and what companies can do and then say, "Well, in a comply and explain regime, we would like to explain to you why we think what we have done complies with what you set out." We do not think there is a strong enough causal link in that area, so we propose strengthening that considerably.

This provides a much stronger foundation with which to address, in particular, harms that arise to adults. At the moment, clause 11 of the Bill, which refers specifically to that, is extremely weak. It is set out in a very odd way that requires platforms only to deal with issues, not to resolve, mitigate or prevent them. It is a clear choice that the Government have made, because they have mitigate and prevent language in the criminal duty and the duty towards children. Clause 11 also suggests that they should only be dealing with things, as Ellen was saying, in their terms and conditions, which is a downstream issue that arises long after you have taken design decisions about how platforms are. So we are tying things much tighter.

We think this broader foundation duty then gives Parliament and the regulator—however that is resolved in the final legislation—the ability to say, "Well, now we can see all the risks of harm that arise, these are the ones we would like you to focus on," based on a much better understanding of what might be happening on the platform.

Q65 **Christina Rees:** Building on that, are any specific gaps in platforms likely response to the new regulatory framework?

William Perrin: The Government have sought, as far as we can tell, to quite rightly allow companies some flexibility in how they comply with guidance from Ofcom. So they have set out essentially a form of comply or explain regime—the sort of thing that comes from the old Cadbury rules around corporate governance in England and Wales—and we feel that some platforms will do that very honestly and openly, and they will do a phenomenal job. I am a great believer in regulation that companies, or the people you are regulating, have the answer to solving the problem. You must not automatically assume that they are trying to be bad. They are just doing what they can to earn returns for shareholders. It is perfectly legitimate for democracies to say there should be different rules that reflect harms to society or returns to society.

Some platforms will come up with a different route to complying than Ofcom had ever conceived of that will be absolutely fine. But our strong suspicion is that some companies will not do that, and they need to be pushed rather harder to redo their risk assessments under guidance from

Ofcom, possibly conforming to some external standards. We think that is a major gap in the regime that we have tried to tighten up. It is okay to take a comply and explain approach, but if you are essentially wilfully blind, because you think that your company is doing such a fantastic thing that the harms that arise from it are actually trivial compared to your mission, or they are so lucrative, given the returns to shareholders, that you just do not want to deal with them, then the regulator needs to be able to tell you that you are going to do another risk assessment, and you are going to do it to some external standards that it sets. We think that is a major gap.

There are a lot of smaller gaps. There is some circularity around having to know what harm is going to arise before you assess the risk. That is a little bit odd. Maybe that can be tidied up with some better drafting. Any harm prevention regime has to revolve around the risk assessment and so you have to make that as strong as possible, or as comprehensive as possible, and then choose what you target within that. That is the particular gap that we have been seeking to tidy up.

Q66 **Christina Rees:** If the Government decide not to introduce an overarching duty of care, are there changes you would like to see to the draft duty on platforms to set out how their terms of service deal with legal but harmful content to ensure platforms are effectively addressing harm from this content?

William Perrin: It needs to be a process that is much more driven by regulatory and civil society interrogation. It must not just be, "So we set out what we are doing our terms and conditions, and that is okay." There needs to be a process that assesses whether that is effective in mitigating or preventing harms—very similar to the one set out for children or for adults. It is similar in principle and indeed more familiar to anyone who has come across statutory duties of care in other areas of law. The Government have chosen quite deliberately to make clause 11 relatively weak. There is substantial pressure for them to strengthen it and we think that is entirely doable.

In western Europe and its associated islands, generally there is regulation of the most powerful sets of media. The European convention on human rights says it is perfectly permissible to regulate broadcasting and to regulate cinema—because that is when it was drafted. We are in an age now where these massive advertising platforms that, to earn returns to their shareholders, are exploiting the things people write and things people say in order to sell more advertising. Now is the time to regulate that where it is causing harm to adults, as we do with TV, radio advertising and film distribution, in the context of these platforms. They are not radio, TV advertising or cinema; they are a different thing with different levels of risks. But we are already doing it for these substantial advertising driven media and there is no case for not doing it here.

Q67 **Martyn Day:** Again to you, William. A lot of the abuse that is harmful to adults will come into the category of abuse that is legal. Unlike with illegal content, the draft Bill does not impose a duty on platforms to actively mitigate those risks. Would you like to see an equivalent duty introduced in the Bill?

William Perrin: The duty that we have set out from Carnegie would have that effect, but we have taken great care to bind it by appropriateness and the context in which it is set. One should not be regulating adult content in the same way one does children's content. No one would disagree with that. But nor should all adult content be regulated in the same way. There will be a huge spectrum from very polite, focused discussion about stamps, through to far more bawdy discussion about football teams or some other thing, or a discussion that is designed to take place after the pub is closed and to be a bit lairy, and a bit entertaining or offensive. But people need to be aware, and they can be made aware through regulation measures, of what they are getting into when they go into such conversations so that they understand what is likely to happen to them—so the reasonably foreseeable harms that might occur—in much the same way as we expect to be warned in other areas of society and commercial life before we engage in some risks.

Once they have entered into that discussion, if it goes wrong and it is not really what they expect, and it becomes far more offensive or harmful to them, there needs to be a process whereby they can get some help, they can get out, they can get protected or they can get some resolution. These very dull systems and processes—I am sure, as a Member of Parliament, this is the kind of thing you encounter every day: a simple malfunctioning of complaints processes where an individual person cannot be heard by a corporation or regulated utility or company—are very much at the crux of this. How can people who are harmed, or have had very bad, unexpected things happen to them, get some resolution of that? But also in these platforms, to what extent can they deploy what we at Carnegie sometimes call user defence tools to ensure that they can manage and constrain the conversation, and that their account or identity on a particular social media service does not become forever tainted, forever abused or forever harmed?

As we have heard sometimes in testimony to the scrutiny Committee—you heard this at the end of the previous discussion—people are being put off using particular platforms. Their voice is being suppressed by the amount of harm and abuse they receive because they have no way of defending themselves effectively on that platform.

Q68 **Martyn Day:** Would you like to see a role for Ofcom in assessing whether the platforms are interpreting their duty to deal with legal but harmful content robustly enough?

William Perrin: Yes, it seems to be pretty straightforward. They are the regulator and they should make that assessment. But it should be based on evidence in a good-faith dialogue with the platform in question. It should be based on Ofcom's general regulatory behaviour to be proportionate, to be good at regulation, to regulate responsibly and also to be broadly balanced, in the way Ofcom has shown itself when taking very difficult decisions about different types of content. For instance, there was the Piers Morgan judgment a couple of months ago. There was wide disagreement about that and that is good—there should be healthy debate about decisions a regulator takes—but it took a tough decision in favour of freedom of speech.

Similarly, it has taken other decisions. There was the case that the Free Speech Union brought against Ofcom to judicially review a decision about a piece of Covid-related programming and Covid disinformation. In that case, the judge was very clear. He said, "Yes, Ofcom has fully robust and effective processes that are compatible with human rights law, and it can take these decisions. You might disagree with the decision, but they have a good process for taking a decision." I think it is well placed to do that. It has a long track record of doing it, but has to apply it in a different context now.

Q69 **Martyn Day:** What changes would need to be made to the draft Bill to allow the framework to enable that?

William Perrin: A number of technical changes, some of which we have set out in our proposals. I think much of the enforcement framework the Government has brought forward is pretty sound. It is quite clearly based on previous regulatory frameworks. Ofcom has strong powers to require information from companies. In the background there is a threat of a criminal offence that could be brought forward if they do not supply information. That is quite a contentious issue at the moment. We wonder a bit about penalties and service denial orders. Are some of these things effective enough, particularly with the largest companies? I think Facebook sits on $60 billion cash at bank. It is just sitting there; it is not earning any money for shareholders. Alphabet sits on over $100 billion cash at bank—I think it is $120 billion. In regulatory microeconomics, you would say, "Well, where are the incentives to respond to fines in those circumstances?", because there is this sort of deadweight money there.

As I say, I have great hopes. In my long experience of regulating all sorts of different industries, companies tend to respond in good faith and the bigger they are, the more they respond. We can see that across a range of sectors, but you have to get the regime right and you have to give the regulator the right powers. As I said earlier, in particular it is this ability to chase through the risk assessment, because the risk assessment is at the heart of this. If you have not been acting to mitigate or prevent the things you have revealed in your risk assessment, that is at the crux of the regime, so we need a stronger ability for Ofcom to drive through a risk assessment and for it to bring in these enforcement powers that it has.

Q70 **Chair:** Just following on from that, in order to undertake these risk assessments, the proposal is that there will be codes of practice in place? I wanted to ask you about this, William, but also Ellen as well, to follow up. How prescriptive do you think these should be? We know one of the issues a lot of people have raised is how fast technology develops and how quickly legislation and regulations can go out of date. What are your thoughts on the Government's proposals and also how we think they could best proceed with codes of practice in particular?

William Perrin: One of the interesting things that Ofcom will do in producing guidance and codes, which has not been studied in great detail, is the production of risk profiles. Ofcom will do its clause 61 risk assessment and then it will produce a series of risk profiles which are essentially, for service types, what is the statistically likely incidence of risk against that service type? And then guidance flows—I think, because it is a bit complicated—from risk assessments that are related to those risk profiles. That is the fulcrum around which it pivots really—the risk profiles. How prescriptive the codes need to be is a balance. We can see this in the draft codes the Government have produced: the children's code, for instance, is really quite prescriptive in many places, but it is based on an acutely victim-led approach to tackling harms, and one should, in good practice, be victim-led. It is quite important to do that and then step back and take a more strategic view than an individual victim perspective and say, "Well, this is the way you tackle those strategically."

Again, I come back to the sense that the biggest companies should respond quite well to some of this. They need to have a dialogue with a regulator that they trust about how they comply with the codes and the drafting of those codes. There is always a risk that companies will just fall back into litigation around the prescription of the code. We saw this in the early days of Ofcom. I did a lot of work on setting up Ofcom nearly 20 years ago now—when I had a full head of hair and a lot more energy—and the first years of operation were often characterised by heavy litigation. There were many, many appeals against its regulatory judgments that bought a lot of Aston Martins for a lot of barristers in chambers. After, I think, eight or nine years of that, Parliament said, "No, we need to limit the rights of appeal because we have to trust Ofcom to get on with it."

Good regulatory practise is the ability to take good decisions quite quickly. There does need to be this ability for the regulator to act and for people to have confidence in the regulator's actions without endless litigation. But we cannot quite tell yet how it will all end up. The more prescriptive your code, in some cases, the more litigation you will get around the detail. Similarly, though, the broader the code and the less prescriptive it is, you might then get issues of definition and fairness, so the balance needs to be struck somewhere in the middle.

Q71 **Christina Rees:** Ellen, obviously say if you have anything to add to that, but I also wanted to ask you about the argument about recognising the more gendered nature of many online harms, including abuse, and how the codes of practice could reflect that. What thoughts you have got on that?

Ellen Judson: On how prescriptive they should be, I think they definitely need to be responsive to actual risks that are arising on specific platforms. I think just mandating that this particular design feature or that particular system is the one that you must have is not going to work to actually reduce risks in practice. It is important that what is prescribed in codes of practise is measurable and something through which it is possible for the regulator to be able to judge whether the platform is complying, beyond just the platform saying that it is complying or producing a top-line transparency report that says, "Yes, we put in place this system so now everything is fine." We want to be able to see if there actually is a result on the levels of harm that people are experiencing.

On that, and related to how Ofcom would be engaging in that kind of regulatory action, the role of independent researchers and civil society is really important here—I declare my conflict of interest as a member a research institute—but I think data access is one of the current gaps in the Bill. I know that Ofcom has significant information gathering powers, which I think is appropriate, and that there is a recognition that there should be more consultation. There should be more thinking about what data access for civil society from social media platforms should look like. It is going to be absolutely crucial as civil society to assist Ofcom in holding platforms to account. Ofcom is only ever going to have finite resources. There is a wide array of expertise that could be brought to assist in that process, but without access to the data—this is something that we face in our own research—there is really a limit to what you can measure, test or observe.

In the case of Facebook and New York University researchers, it shut them down for carrying out their research because it went against Facebook's terms of service, I believe. If there are cases where there is a clear public interest in that data being available and it is needed in order to test how well the codes of practice are being implemented, if they are being implemented, how well they are working and what the effect is, I think having stronger mechanisms for being able to do that—obviously taking data privacy into account—would be really useful.

On the point about gender, I think what it is going to be really crucial is that specificity of responding to the actual risks. What the risk of gendered abuse and gendered disinformation look like, and what the solutions are, are going to vary hugely across different platforms and different contexts. As Bertie was saying earlier, in different political contexts in response to different political issues, just using the same tool kits is not going to be very effective. Thinking about gender disinformation campaigns, we know that they evolve in response to action taken against them, so if the algorithm is able to detect certain gendered slurs, people will change the characters in words so that they are able to fly under the radar. Always trying to play catch-up with these sorts of campaigns and these sorts of widespread problems of gendered abuse at scale is always going to be limited.

That is why some of the elements which were being discussed earlier around better digital citizenship, better digital literacy, and empowering users to be able to control their own online experiences are really crucial, but those are all ways users can protect themselves. We also need to be thinking about what platforms are responsible for. What is their role in this? Where are they making decisions that may be increasing how gendered abuse is spreading, or normalising gendered abuse as just the way that people communicate on that particular platform? On the flip side of that, what are the platforms where gendered abuse is perhaps less of an issue? What are they doing right, what can we learn from them and how can we apply that to the expectations that we have of the platforms where it is a problem? I think all that requires measurement and investigation, and we would hope that the regulator would have the powers under the Bill to be doing that kind of granular investigation of saying, "We see that there is a problem. We have heard"—whether anecdotally or through research—"that this is a very significant problem. Tell us what you are doing about it and provide the actual evidence of what impact it has had on the amplification or the spread of gendered abuse."

Q72 **Christina Rees:** Bertie, we will come back to anonymity now. Previous witnesses have suggested that a significant proportion of online abuse is posted anonymously and that being able to post anonymously encourages abusive behaviour. Has anonymity come up as an important factor in your research on the scale of hate speech online?

Dr Vidgen: Anonymity is one of these really difficult issues—I think there is widespread consensus around this—because both vulnerable and non-vulnerable people who want to be anonymous online, and it is incredibly important that we protect that. Whatever finding we have around the relationship between anonymity and abuse or hate, we have to take into account that anonymity is very important. We heard in the previous session about how Twitter defines anonymity, which is very interesting. There is a very clear body of research and theory, in particular, which says—

Q73 **Chair:** Just on that, how do you define anonymity? When you say anonymity is important, do you mean anonymity as in having an account that is not verified in any way or that you are untraceable, or do you mean anonymity as in you can present an argument and people do not necessarily know who you are, but the platform, or a third-party provider, would be able to trace you if you were being abusive? Have you come to conclusions on those definitions?

Dr Vidgen: You have taken the word that I was going to use from me, which is "traceability". This is the key distinction: are you anonymous or are you untraceable? Traceability is crucial for security services. If we want to have security services that can prosecute illegal forms of abuse, absolutely we need people to be traceable. I think that is really what Twitter was talking about. It is saying, "Look, we have a way of tracing these people." It is not that robust really, if all you need to do is have an email address and a phone number. As we all know, that is not a big barrier.

Anonymity, of course, is slightly different, and if we talk about it in a very technical sense, there are certain things around how we define anonymity, but in a social sense, I think we just mean: can you hide who you are online? There is also the issue of people who have pseudonymous accounts—someone pretending to be someone else, or even when it is slightly recognised that the account is not completely authentic. These are slightly different issues. But there is that body of research that says, yes, anonymity is going to be absolutely crucial because essentially you are no longer responsible for your actions. When you add to the fact that you are in a computer mediated environment, you do not see the immediate effect of whatever you say on the person who is affected by it, so you can be incredibly abusive and you do not see how much it is harming them. It is often asynchronous, so you can post late at night and forget about it, but for the person who has received it, of course, it has a huge impact.

There is a lot of what I would call theory-inspired research which says that this should be the case, and a huge amount of qualitative studies show this is the case. I have not seen as much quantitative, systematic, longitudinal research that shows that anonymity is the key problem that we need to solve. It absolutely is part of the problem, but I think some people present it as the singular issue, saying, "If we can just solve anonymity, we solve online abuse." I definitely do not see that in the research that we have done. Lots of abuse online can be identified to a person who has to bear the consequences of that, and yet they still do it and they do not necessarily even see it as abusive. They think they are just letting off steam or it is valid for them to say something which is hateful and prejudiced. This goes against some of the general discourse, which is certainly leaning towards the idea that anonymity is the key problem. We do not see that in all the research we have reviewed, though certainly it is part of the issue.

I think the other thing with anonymity is thinking about safe by design systems. So it is absolutely right that people can be anonymous online—that is really important—but there should also be a way of designing systems so that if I do not want to see content from people who are anonymous, I should have a way of doing that because that might help me to limit my exposure to not just abusive content, but disinformation and other forms of harm. I think that is really where the conversation should be heading towards: how do we have systems that give people that choice and flexibility, rather than trying to be too prescriptive about anonymity as a single issue?

Q74 **Chair:** Do you mind if I just follow up on that with two things? First, when you say people should have the right to be anonymous online, there is the counter argument that many members of the public feel that they do not have the right to be anonymous walking down the street, in a public meeting voicing an opinion, or in a shop if they shout at somebody, so they would question why you have the right to be anonymous online. That is just a question that has been put to us as part of undertaking this inquiry.

The other question I would put is: if there is no research to show that anonymity is the cause, if you like, of all the online abuse, has there been any research to show what is the cause of the predominance of abuse online, which is much more amplified than in the physical world?

Dr Vidgen: I just want to make a small clarification. I am not saying that there is no evidence that anonymity is important; it is just whether people are overstating the importance of anonymity, and sometimes it is treated as the single factor. That is the thing that I have an issue with. There is evidence that shows that it is one factor among others. Why is anonymity important? It can range from people who are dissidents or journalists, and who are speaking out politically—obviously that is incredibly sensitive, especially when you look globally—and that needs to be protected. But it goes all the way through to people who are exploring their identity or searching for a sense of self. They might benefit from anonymity or having pseudonymous accounts as well. There are some very clear motivations for having anonymity protected.

In terms of the reasons why, we might ask, "What are the causes of online hate and online abuse?" I think Seyi said it really well in the previous session: it is often a continuation of what we see in the offline world. There is a lot of research around how people who only really interact with people from their own social group can have a real lack of understanding of and contact with people from other groups, and can start to develop some very prejudicial and hateful views about them because of that lack of contact and engagement. That spills over into the online world as well. There are also these issues around computer mediated communications and the sense of separation from the person who bears the brunt of what you say. There are other factors, too, but the real challenge is that social media is changing so rapidly, and it is so new, that our ability to generate longitudinal, systematic, quantitative research into the causal drivers of this sort of behaviour is very limited. As we just heard, one of the big challenges is how you get the right data. All this data is not held by us as researchers; it is held by the platforms. If we cannot access that, it is very difficult to say, "Look, this is the problem. Now let's go solve it." We are still just trying to tell you what the problem is, and we will not be able to do that fully until we can get better data.

Ellen Judson: Might I just respond to some of that as well? On the question about the right to be anonymous, I think there are lots of cases in the offline world where we accept a degree of the right to be traced. When we go to the shop, for instance, we know that there are probably going to be CCTV cameras, but we are not presenting an ID to the person at the front door so that they can definitely track us down if we happen to commit abuse while we are in that space. Over the past few years, we have seen a lot of debate around where the limit is between how trackable and traceable—to use a common term—it is acceptable to be?" We saw Google and Apple refusing to make changes, for privacy reasons, to some of the Covid apps that would have made people more traceable. It is a live debate, even in the offline world, of where the limits of traceability are.

I totally agree with what Bertie has said about anonymity often coming up as, "Maybe this is just the solution, and if we fix this, everything will go away" and not thinking that is going to work. There is not a reduction of anonymity that only negatively impacts people who are abusive. That is often the silver bullet people are looking for, quite understandably. But anything that is reducing people's ability to be anonymous, whether it is directly by requiring identity or by proxy measures which strongly incentivise them to reveal their identity, is going to put some people at risk.

A worry that I have about a system that allows people to block all anonymous accounts, for instance, regardless of the actual behaviour of those accounts, simply on identity basis, it is not connected to whether that person has been abusive or is spreading disinformation; it is connected simply to the fact that they are not revealing their identity, which we know is disproportionately going to be marginalised groups. Then the risk is that this leads to a two-tier system where we say, "Well, people who have to be anonymous for their own safety are allowed onto the platform, but they cannot talk to anybody and they cannot interact in public debates because we are not entirely sure who they are."

I think the distinction between people who use pseudonyms but provide IDs, say to a platform, and people who do not provide IDs to a platform is a useful distinction. But also we should keep in mind that people who need to use pseudonyms online are often maybe also unable to give their personal verification data to a platform, and that could be for practical reasons. Perhaps they do not have a bank account and they cannot use the third party identity verification that links to the bank account. Or it could be for safety reasons: the risk of leaks or hacks, or of bad actors having access to that data or platforms misusing that data—we have seen that platforms can do that—is just too high for them to take that risk. I think, particularly in the context of a Bill which is so focused on safety, one of the things we would really like to see is a greater recognition of how privacy and anonymity online do entail safety for some users.

The Bill at the moment does not speak about anonymity specifically, but things about privacy and freedom of expression were brought in as constraints: "We are going to pursue safety, but while you are pursuing safety, you also have a duty in regard to the importance of these other rights." Carnegie's proposals for reworking the Bill do this nicely. Actually, for many people, being able to exercise those rights and the pursuit of them is absolutely central to them being able to be safe online and exist and speak safely in the online world.

Q75 **Chair:** I was just thinking back to something that Bertie said earlier: we do not have the research to show that anonymity is the root cause of all the abuse. I think the phrase you used was that some of it is computer automation and that distance. What you said about the need for more research and more data available was interesting. If we are going to solve this, you have got to have the data to show what is causing it. I think that has come across very clearly from what you are saying.

You have proposed a few solutions there, Ellen, Is there anything else that you would say about what possible solutions there are to increase traceability to enable the widest participation possible, but minimise the ability to abuse? I know not all abuse is anonymous, but that, I think, is the most egregious abuse for people—when they do not even know who is abusing them. I think that is one of the big challenges.

I am conscious that we are running out of time. Do any of you have anything you would like to add that you do not feel you have had the opportunity to address so far? We have got a couple of minutes, really.

Dr Vidgen: Yes, just two final points. First, I think that the people who are most affected by abuse should be the ones who have a much bigger voice and a much bigger role in the whole policy discourse and civic discourse around this problem. People who have been affected, and groups who are more likely to be affected by abuse, should be centre stage. That should be key across all of the work we do.

The second point, really, is just clarity. There is so much confusion around what we mean by any of these terms. If we have more precision and clarity around what we mean, then even if I disagree with your account of abuse, I know what it is and I can challenge it. That would be incredibly helpful. And maybe just one request for this Committee is that if you could make any push towards having clearer definitions, and more consensus and standardisation, that would be a huge help.

William Perrin: One thing I was particularly struck by, Chair—I read this on your website as well—was the very different views of school children, in particular, in that particular example about anonymity. It is symptomatic of the transition from the first era of the internet, which was about individual personal freedoms: "Governments are not welcome here. Corporations are not very welcome here either. This is a free and open internet." We are now transiting from that towards what Jonathan Zittrain calls a more public health era of the internet, where in fact we are dealing with the externalities of all the individual freedoms which, taken together, cause harms to society. This is a strong role for the regulator in its research. Ofcom, with its broadcasting responsibility, has a very strong track record in qualitative and quantitative research—so does the BBFC, in regulating film distribution—on public attitudes and understanding. Some of it is so detailed, particularly the rank list of swear words that changes every now, that it is almost comedic. But it does greatly inform what is an acceptable level of risk trade-off in society. The very action of the regulator being able to conduct that research and also to work with the companies to understand their research better, and to bring that out into a regulatory space where they can discuss it, will greatly improve everybody's understanding of how to prevent online harm and abuse.

What you have also seen, I think, and what other Committees examining these issues have seen, is the astonishing and quite distressing breadth of issues that arise. The Bill does need to be broad enough to understand all those things and then decide where to focus within some democratic framework, informed by the research I was talking about earlier. At the moment, the Government's Bill is a little bit too tightly focused on constraining a regulator rather than taking that broader view first and then focusing hard. We will see what the Joint Committee recommends and what your Committee recommends. I am sure when the Joint Committee’s report is debated, you will bring a strong voice to the Floor of the House on that. I hope the Secretary of State who, I thought, showed a commendable spirit of team working to the Joint Committee, will extend that to the output of your work as well and amend the Bill appropriately.

Q76 **Chair:** Ellen, a final word.

Ellen Judson: Just in response to your question on anonymity. I think what is really important is to separate and identify exactly which problem which intervention is trying to solve. We have the problem, as you said, of what happens if someone is abusive online and committing an offence online and we cannot trace them. I think that is much less of a question about the Online Safety Bill and what it is concerned with, and more a question about the resourcing of law enforcement and how they are conducting these investigations. How are they supporting people who have been targeted in this way? Are social media companies handing over the data that they are already required to through existing powers?

There is a separate question: how do we reduce the risk of abuse happening in the first place? I think focusing on traceability risks misses that preventative piece, because accountability is going to always be post hoc, but it is also only really going to come into play when we are talking about illegal abuse, and so much of the stuff that we see online probably counts as legal speech. Without that kind of accountability framework, you are then left in the dark as to how to tackle it. It is focusing on how we can prevent abuse from anonymous accounts and non-anonymous accounts. As to what that looks like, again, as we have said, a lot of it depends on specific platforms, specific risks and specific evidence. But it would be really useful for the regulator to be investigating that sort of thing, as would it asking platforms for their evidence as to what they have found actually helps to reduce those risks.

Just a final point. As I said, we would very much like to see the understanding of safety being broadened out within the Bill to understand how safety looks different for different groups, and different rights and freedoms need to be a bit more prioritised. I think it is also good to keep in mind that this is going to be world-leading regulations. Obviously we have got the DSA and other countries have taken some steps, but I think a lot of countries will be watching to see what the UK does. That is why it is particularly important that rights and freedoms are prioritised explicitly to avoid copycat regulation in countries with different frameworks where rights are not embedded into everyday practices. I would like to see that considered in a redrafted Bill.

Chair: Thank you so much for your time today.