29th September, 2021

Yvette Cooper MP,
Home Affairs Committee,
House of Commons,
London, SW1A 0AA.

Dear Ms. Cooper,

Thank you for inviting me to give evidence to the Home Affairs Select Committee. Please find below our response to the topics Members of the Committee sought further information on.

Q245: During the evidence session, the Chair sent to you examples of racist abuse sent to the England players in July that were still visible on 8 September – 8 weeks after the Euros final. Later on the same day, you emailed Committee staff to confirm that all the offending posts identified by the Chair had been removed. Please explain why these posts were online after such a long period of time, and why these very clear examples of racist abuse were not detected?

Across Facebook and Instagram we have strong policies against hate speech and racism and have significantly improved our technology to help us reduce the prevalence of violating content. However, zero tolerance does not mean zero incidence. In relation to the horrendous abuse we saw on the accounts of the England team players following the Euros, we removed 15,000 comments from 11-13th July for Bullying and Harassment and Hate Speech directed at members of the England team, the majority of which were found and flagged proactively using our AI technology. The majority of these comments were also removed within an hour of them being posted. However, the Committee subsequently raised 5 pieces of content on 8th September, all of which we removed within minutes of investigating, as well as disabling several accounts for violating our policies.

As discussed at the hearing and in my earlier letter to you, AI detection and indeed human review are not perfect and we are sorry to see these comments on Instagram a few weeks after being posted. When it comes to hate speech, context is extremely important. Words, emojis and phrases, on their own, may not be inherently violating – often we need to see the context in which they are being used, as well as who is sending them and to whom, and AI alone does not always get this right. We also need to understand whether the words or emojis that may be used to abuse someone, are being used to condemn hate speech, and whether certain words

that have been used to racially abuse are being used as reclaimed speech. AI technology and human review are therefore both important, but so too is in-app reporting.

At the hearing you mentioned the potential for people to see the racist abuse and the impact that may have and I want to provide some reassurance on that point. Our systems seek to automatically detect and downrank comments that may violate our policies. So those viewing accounts and replying to posts from the footballers would be less likely to see the comments posted that may be offensive or racist. In addition, while we do not want any comments on our services which violate our policies, these 15,000 comments represented a small proportion of the hundreds of thousands of positive comments and millions of likes that were posted on the players' accounts.

We have strong rules against [hate speech](#) and racism and we use a combination of Artificial Intelligence and human review to identify and remove content that violates our [Community Guidelines](#). We have made significant progress in our ability to proactively detect and take action on such content. For example, in April to June this year, 95.1% of the 9.8m pieces of hate speech content we took action on on Instagram were found and flagged by us before anyone reported them. This is a significant increase in our proactive rate from 42.5% just over a year ago in the first quarter of 2020. We've provided a detailed overview of the technological improvements that have driven this improvement regarding hate speech detection [here](#), and we provided a further update on our enforcement [this month](#).

Perhaps most importantly, we are determined to work with others – including footballers and the football community – to stop hate from being posted in the first place. Removing content that has been posted already is just one part of the picture, especially as we seek to protect people from having to experience abuse while continuing to reach fans across the globe. This year we launched:
- A stronger comment warning discouraging people from posting something hurtful in the first place (in a given week, potentially hurtful comments were edited or deleted about 50% of the time after seeing these warnings);
- A tool to filter comments and Direct Message that contain offensive words, phrases and emojis, so you never have to see them (we're working closely with football bodies to help players turn this tool on); and
- A tool to automatically hide comments and DM requests from people who don't follow you, or who only recently followed you, during moments when you expect an increase in comments and messages. Our research shows that a lot of negativity towards public figures comes from these groups, including after the recent Euros final, so we believe this feature could have a meaningful impact in reducing the amount of abuse public figures experience.

All of this is in addition to the very important work that we and others do to make clear that hate is unacceptable, online and offline, such as our Take a Stand initiative with Kick It Out.

Q248-250: Precisely how many senior people do you have designing and developing your AI technology and making decisions on this technology who have themselves experienced racist

abuse? In addition, please specify how many people working for Facebook (and Instagram) – in particular, senior people – are from Black, Asian and minority ethnic backgrounds. By senior people we mean people employed in management roles overseeing an area or department including, for example, Heads of areas (like public policy), directors and executives.

Our CEO has been clear that diversity in leadership positions and across the company is a priority for him personally and the rest of the leadership team at Facebook. Connecting the world takes people with different backgrounds and points of view to build products that work better for everyone. This means building a workforce that reflects the diversity of the people we serve, including people's lived experiences as set out in the Committee's questions.

The Committee asked about the makeup of our AI and product teams, and the ways in which people's lived experience informs our work. While we do not collect and share data on the specific experiences you referenced, to build products, develop policies, and best serve people in our global community, we must apply diverse perspectives to everything we do. We leverage the diversity of the people who work at Facebook to provide strategic input on our products, policies, programs, and practices and apply these learnings to build for all. We created Inclusive Product Councils that offer diverse perspectives and feedback to product teams across the company based on lived experiences. Input from an Inclusive Product Council reduces the risk of harm and helps our teams consider how a product may impact a diverse range of communities and people. Participation in an Inclusive Product Council is encouraged and recognised in our performance evaluation process. When it comes to ensuring fairness in AI more broadly, we believe AI can work well for everyone, and we've outlined in detail here and here how we are tackling some of the hard problems to help get there. This includes further information on our interdisciplinary Responsible AI team and 'Fairness Flow', our technical toolkit that enables our teams to analyse how some types of AI models and labels perform across different groups.

Since 2014, we've also publicly reported Facebook's diversity metrics and shared our plans to better support communities of color, women, members of the LGBTQ+ community and others. And over the past two years, we set three goals to increase representation in our workforce over five years:
1. Double the number of women employees globally and double the number of Black and Hispanic employees in the US.
2. Increase the number of people from underrepresented groups. By 2024, our goal is to have at least 50% of our workforce comprised of women globally, and underrepresented minorities, people with two or more ethnicities, people with disabilities and veterans in the US.
3. Increase the number of US-based leaders (Director-level employees and above) who are people of color by 30%.

Our progress against these goals and our latest diversity report can be found here. It shows that over the last year, we've seen growth in underrepresented communities at Facebook and today, more than 21% of our non-technical employees identify as Black or Hispanic. We're also pleased to report that we had our most diverse intern class ever in 2021, with 44.0% women globally

FACEBOOK

and 20.4% from underrepresented minority communities in the US (Black and Latinx). This year, 4.7% of our US-based employees identify as people with disabilities, and 2.4% identify as veterans. Our LGBTQ+ community makes up 10.6% of our US-based workforce.

The Committee asked in particular about representation at leadership level. On the third goal we're pleased that in just one year we achieved a 38.2% increase in Black leaders due to strong recruiting and increased focus on retaining top talent across the company. As the company grows, we'll work to maintain this representation of Black employees in leadership. Further detail on the representation of our leadership team can be found in our latest report. We will also continue striving for an overall 30% increase in representation of people of color in the US, including Asian and Hispanic people, in leadership roles.

More broadly when it comes to company leadership, Maxine Williams, our Chief Diversity Officer, reports into Sheryl Sandberg our COO and forms part of our leadership conversations and decisions. And her role means we are placing a structural focus on more inclusive products and policies and elevating diversity and inclusion in all management team discussions. This year we also appointed Roy L. Austin, Jr., a renowned civil rights attorney and advocate, as VP of Civil Rights and Deputy General Counsel in order to establish the company's new civil rights organisation.

Finally, we are committed to supporting the next generation of talent. We've continued to invest in education programs designed to increase access to STEM and computer science training for women, Black and Latinx people. This year, we're expanding our co-teaching AI program that started as a pilot in 2020 with Georgia Tech. Together with the university, we built a deep-learning curriculum that professors at Historically Black Colleges and Universities and Hispanic-Serving Institutions will adopt. Our goal is to enable more students from underrepresented minority communities to be trained in AI. We recently announced a new Summer News Fellowship for undergraduate students and recent graduates of Historically Black Colleges and Universities. This program will elevate emerging Black talent and future media professionals by pairing them with established national and local news organisations.

These are in addition to existing partnerships and programs we have with CodePath.org, Align, TechPrep and other organisations dedicated to levelling the playing field. These programs help underrepresented minority students who face disproportionate challenges gain equitable access and opportunity to pursue degrees, and eventually careers, in computer science.

I hope this information is of assistance to the Committee.

Yours sincerely,

Tara Hopkins
Director Public Policy EMEA, Instagram

FACEBOOK