

## Written evidence submitted by Facebook and Instagram (MISS0039)

### Introduction

The Facebook Company (Facebook) was built to help people stay connected. Our mission is to give people the power to build community and bring the world closer together. We're committed to building technologies that enable the best of what people can do together. Our products empower more than 2 billion people around the world to keep in touch, share ideas, offer support and make a difference. Over \$2 billion has been raised by our community to support the causes they care about, over 140 million businesses use our apps to connect with customers and grow, over 100 billion messages are shared every day to help people stay close even when they are far apart and over 1 billion stories are shared every day to help people express themselves and connect.

We view the safety of the people who use our platforms as our most important responsibility and we have developed robust policies, tools and resources to keep people safe. We work closely with external safety experts from across the world who help inform our policies and practices on safety, as well as with industry and community partners to understand their experiences on our platforms.

We work closely with trade bodies, regulators and partners, supporting industry-led initiatives to share knowledge, build consensus and work towards making our online services as safe as possible. We have said [for some time](#) that we would welcome a more active role for Government, and that if designed well, new frameworks for regulating harmful content can contribute to the internet's continued success by articulating clear ways for government, companies and civil society to share responsibilities and work together. Keeping people safe online is complex, requiring well designed regulation that protects people from harm without stifling expression and the benefits that come from online connection. We reflected this view in our response to the UK Online Harms whitepaper.

### **1.1 Facebook's Community Standards and Instagram's Community Guidelines govern the content that we do and don't allow across Facebook and Instagram, and are designed to ensure our communities across both platforms feel free and safe to express themselves.**

Every community has standards, and since Facebook's earliest days it has also had its [Community Standards](#) – the rules that determine what content stays up and what comes down on Facebook. We always try to strike the balance between giving people a voice and place to express themselves while preventing real world harm and ensuring the safety of the people who use our platforms.

Instagram's [Community Guidelines](#) are designed to foster an authentic and safe place for inspiration and expression, while encouraging our community to respect each other and their diversity of perspectives, beliefs and cultures.

The team responsible for setting these policies is global -- based in more than 10 offices across six countries, to reflect the different cultural diversity of our community. We have more than 35,000 people working on safety and security for both Facebook and Instagram, including content reviewers who are native speakers of every language widely used in the world. We have

offices in many time zones to ensure we can respond to reports twenty four hours a day and seven days a week, and we invest heavily in training and support for every person and team. To be transparent about the progress we are making against harmful content, we issue a regular transparency report (previously bi-annual, now quarterly), which includes a [Community Standards Enforcement Report](#) that shares metrics on how we are doing at preventing and taking action on content that goes against our Community Standards. For example, from January to March 2020 we took action on 1.3 million pieces of suicide and self-harm content (the category into which eating disorders fall) from Instagram, and 1.7 million pieces of content from Facebook.

In April 2018, we [published our internal guidelines](#) that our teams use to enforce these standards, to help provide more transparency around our policies and where we draw the line on complex policy areas. These guidelines are designed to reduce subjectivity and ensure that decisions made by reviewers are as consistent as possible. Our policy process involves regularly getting input from outside experts and organisations, such as our [expert advisory group](#), made up of mental health organizations and academics from more than 20 countries, to ensure we understand the different perspectives that exist as we balance giving people a voice with safety, as well as the impacts of our policies on different communities globally. We regularly review our policies to ensure they keep up with the way people are using our features, and new academic research and insights, and every few weeks the team runs a meeting to discuss potential changes to our policies based on new research or data. For each change the team gets outside input -- and we've also invited academics and journalists to join these meetings to understand this process. We [publish minutes](#) from these meetings, and we plan to include a change log so that people can track updates to our Community Standards over time.

## **1.2 Both Facebook and Instagram have specific rules intended to combat content that may lead to harm in relation to body image.**

We understand we have a big responsibility, particularly to the young people who use our products and vulnerable people who may be experiencing mental health challenges, including eating disorders.

Mental health is complex and affects people in different ways. Some people find it helpful to share their experiences and get support from friends, family and others in the community. Social media can also help tackle stigma associated with mental health. At the same time, experts from our Suicide and Self Injury Advisory Board (whose UK members include academics from Sheffield University and the Samaritans), tell us what's helpful for some may be harmful for others. Finding the right balance is really important - we need to protect vulnerable and impressionable people from content that could put them at risk - and we also want people to be able to express themselves, seek support, find a community and tell their recovery story.

Our policies do not allow content that promotes or encourages eating disorders. We've developed these policies in consultation with a global advisory group of experts, including the eating disorder charity, Beat, in the United Kingdom (UK).

Under these policies, we do not allow graphic content around eating disorders, or any content that encourages or promotes eating disorders. We also do not allow content that focuses on the

depiction of ribs, collar bones, thigh gaps, hips, concave stomachs, or protruding spines or scapula when shared together with terms associated with eating disorders. We remove any of this content as soon as we identify it. We also remove content that contains instructions for drastic and unhealthy weight loss, when shared with terms associated with eating disorders.

However, we don't remove all content discussing eating disorders because experts tell us that people sharing their experiences of mental health, including eating disorders, can play an important part in supporting people's recovery. It could also be a cry for help, and removing it may make the person posting feel more isolated, and deprive friends and family of the ability to see and reach out.

For example, we might allow people to share images of healed cuts or self-harm imagery, to discuss their journey to recovery or offer moral support for other people who might be struggling with dark thoughts and looking for support. We might also allow the depiction of content on Facebook and Instagram that focuses on the depiction of ribs, collar bones, thigh gaps, hips, concave stomachs, or protruding spines or scapula that is clearly in a recovery context.

This doesn't mean we have no safeguards for this kind of content. While we may not remove this content, we add a sensitivity screen, with a warning that the content may be upsetting to some. We also add an additional moment of friction when someone wishes to reshare the content by sending it to someone or posting it to their Stories, reminding them that this content may be upsetting to some and asking if they are sure about sharing it. By doing this, we walk the line between destigmatising mental health discussions and protecting people from content that may otherwise be triggering for them.

We also prevent content related to eating disorders from appearing in parts of Instagram where we recommend content to people, such as Explore. Once a hashtag is identified as sensitive it cannot be recommended in-app via Explore or in Feed and it will not appear on a related hashtags list. When people search for hashtags that may involve sensitive content, we ask them if they need any resources and direct them to local organisations and resources that could help. While there may be good reasons for people to be searching for terms related to an eating disorder (such as recovery or finding support through a difficult time), we will still not allow content within the hashtag page that actively glorifies or promotes such eating disorders.

We have also developed a tool which uses machine learning to automatically identify hashtags that are being used to share eating disorder content which will remove them or add a sensitive content screen before users view the related content- for example hashtags that include 'pro-ana' or pro anorexia phrasing. That technology learns over time to recognise variations of a hashtag once it is banned to allow us to take the appropriate action. If we find that certain hashtags are consistently being abused or if the hashtag itself inherently promotes self-harm and eating disorders in any way, we will make it unsearchable.

We encourage anyone who comes across content that makes them feel uncomfortable or they believe to be in violation of our Community Guidelines or Community Standards, to report it using our in-app reporting tools. We prioritise all reports related to eating disorders, self-harm or suicide.

Eating disorders and mental health are complex issues, and we recognise that as technology experts, it's important that we work with people who are experts in these fields. That's why we have developed long-standing relationships with leading mental health organisations to help us to better understand eating disorders and to help ensure we have the right policies in place. Our engagement with experts, such as Beat in the UK, has proven so valuable that in addition to the teams that are dedicated to wellbeing in the product development, and in our global safety teams, we have hired experts across our teams to focus on health and well-being issues and their impact on our apps and policies, including Facebook and Instagram. The safety team is always exploring new ways to improve support for our community. We also continue to work with our advisory group of mental health experts to inform and guide our policies and practices around mental health.

### **1.3 We take a range of different actions in response to this kind of content.**

We take action against content that goes against our Community Standards or Community Guidelines - this may include removing content or in other instances covering content with a warning screen, to let people know that the content may be sensitive. We prioritise for review by human moderators the content that has the greatest potential to harm our community, including that relating to suicide, self-harm, and eating disorders.

As mentioned, content that encourages or promotes eating disorders is removed as soon as we identify it. This also applies to content that focuses on the depiction of ribs, collar bones, thigh gaps, hips, concave stomachs, or protruding spines or scapula when shared together with terms associated with eating disorders. We also remove content that contains instructions for drastic and unhealthy weight loss, when shared with terms associated with eating disorders.

Artificial intelligence is the best way to quickly and effectively review the amount of content posted on Facebook and Instagram. Some categories of harmful content are easier for AI to identify, and in others it will take more work to develop the technology. For example, visual problems, like identifying nudity, are often easier for AI than nuanced linguistic challenges, like hate speech. We enforce our policies using artificial intelligence technology that proactively identifies content related to suicide and self-harm such as graphic self-harm which violates our policies. We're also working with European regulators to try and bring additional, more sophisticated technology to the EU that we already use in countries outside of Europe that can help us find and act on more content related to suicide, self-harm and, soon, eating disorders. This action could include removing it, stopping permitted content being recommended, and directing more people to organisations that can help.

However, as mentioned previously, we understand that enabling people to discuss their mental health journey or connect with others who have battled similar issues can be an important part of recovery. We have partnerships with 40 helpline organisations around the world. In the UK, people are directed to support from CALM, the Samaritans, and HOPELine UK run by Papyrus, when they post or search for content related to eating disorders. In the summer of 2019, Instagram worked with Young Minds, a UK based charity, to run feedback sessions with young people aged between 14 - 25 who had dealt with and come through an eating disorder or self harming behaviour. We heard directly from these young people about their experiences on the platform and what we could do better to protect them while using Instagram. These organisations

have also helped us with advice on our help centre, to make sure that we're using the most appropriate language to discuss these issues.

Not all of this work can and should be done online, which is why we also work with partners to help people manage their wellbeing and safety offline. We have worked with the US based National Eating Disorders Association to produce in-app content for Instagram and advice on our Instagram Help Centre for anyone with an eating disorder, or for anyone worried about a friend.

In the United Kingdom, Facebook supports Be Real, a national campaign made up of individuals, schools, businesses, charities and public bodies that seeks to change attitudes to body image. It is also a contributor to the Media Smart media literacy programme for 7 - 16 year olds, which provides free educational resources for teachers, parents and young people - including a dedicated body image programme for boys and girls.

#### **1.4 We also view our Advertising Policies as a key tool in keeping users safe from inappropriate or harmful content**

We aim to ensure users are kept safe from advertising that may be inappropriate or harmful, through a suite of policies, tools, content reporting capabilities and resources. Our standards are applied effectively and enforced so that consumers have limited exposure to harmful or misleading advertising.

We have a set of advertising policies that are designed to help keep our community safe. Among these policies are several measures specifically designed to address issues around negative body image in ads, particularly among young people. Ads must not contain "before-and-after" images or images that contain unexpected or unlikely results. Ad content must also not imply or attempt to generate negative self-perception in order to promote diet, weight loss, or other health related products.

We have restrictions on the goods and services that can be promoted, prohibiting unsafe supplements and restricting weight loss products and plans. Certain products or cosmetic procedures can only be promoted to over-18s. We also introduced new guidelines in 2019 to restrict anything that makes miraculous claims about certain diet or weight loss products, consulting with iWeigh founder and actress Jameela Jamil as we developed this new policy.

We also have policies to manage branded content, i.e. organic content posted by influencers or creators in exchange for compensation from a brand, to make clear to users that a financial relationship is involved. Content must tag the featured third party product, brand, or business partner, and this type of content can only be posted to certain Facebook Pages and profiles and Instagram accounts. Creators cannot accept anything of value to post content that does not feature themselves or that they were not involved in creating.

Users consistently tell us that if they are going to see ads, they want them to be relevant. In order to do this, we need to understand their interests, so based on what pages people like, what they click on, amongst other signals, we create categories and charge advertisers to show ads to that category. Although advertising to specific groups existed well before the internet, online advertising allows much more precise targeting and therefore more-relevant ads.

However, we also think it's important for users to understand why they see certain things. On both Facebook and Instagram users can find out why they're seeing an ad and change their preferences to get ads they are interested in. If at any time users feel uncomfortable about the ads/content they see, we have provided a number of tools to further control content, for example: on every ad they see, a user can choose to hide that ad, hide all ads from that advertiser, or report a problem with the ad. If an ad that's already running is reported by our community, the ad may be reviewed again and if it's found to be in violation of our policies, we'll stop running it. The data that we hold about our users stays private.

Responsible advertisers want to uphold the integrity of their brand and, as such, demand a safe environment to advertise. We therefore embed safety into the development process of our products and ensure it is a key pillar of our global strategy. Furthermore, advertisers yield considerable influence and have a key role to play in ensuring we do as much as we can to keep our platforms and products safe because they have the power to cut budgets, thereby playing their role in the self-regulatory system.

Before ads go live on Facebook or Instagram, they are subject to Facebook's ad review system, which relies primarily on automated review tools to detect keywords, images, and a host of other signals that may indicate a violation of one of the Advertising Policies. If this process detects a violation of these policies, it will reject the ad. We use human reviewers to improve and train our automated systems, and in some cases, review specific ads. No such system is—or ever can be—perfect, so we also rely on user feedback and reports from regulators to help us identify possible policy-violating content.

### **Body image in the media**

At Instagram we seek to foster a diverse community which reflects the world around us. We do this by partnering with creators and non-governmental organisations (NGOs), and by helping to give voice to communities who may not find that opportunity in other ways. Our goal is for Instagram to be a kind and supportive platform.

We have partnered with a wide range of body positive creators and organisations, including:

- Instagram and I\_Weigh have developed a six part interview series, hosted by Jameela Jamil, which explores identity and life positivity with every episode featuring body positivity conversations.
- The @instagram account is one of most followed social media accounts globally with 325 million followers. We have used the power of this account to feature voices such as [Nikki Lilly](#) and [Megan Crabbe](#) was featured during the UK Anti-bullying week to promote and support body positive conversations.
- We have partnered with body positivity creator [Chessie King](#) on the Instagram *Create Don't Hate* campaign and the bullying vs [banter Cybersmile campaign](#) which took place during UK Anti-bullying Week in November 2019.
- Off platform Instagram hosted a panel at the UK's largest Video Convention ([VidCon](#)) with body positivity influencer [Sonny Turner](#).

- Instagram is also about to launch a 6-part series with [Megan Crabbe](#) on IGTV which features Megan in conversation with her sister, who has autism, to bring diverse voices and perspectives to the platform.

## **2.1 Research on effects of social media content on body image**

In April 2019, we [joined the Samaritans](#) and other social media and internet companies to partner on industry-wide research to inform our policies and create industry guidelines on suicide and self harm. We have also created research grants to fund additional research that can help us better understand the ways Instagram is used, and how we can better support our community's wellbeing.

In 2019, Instagram requested research proposals to help us better understand experiences on Instagram that foster or harm the wellbeing and safety of our communities and societies. This includes, but is not limited to, research that will help us understand problematic issues facing our communities, develop better policies, assess possible interventions to protect our communities, or identify the mechanisms (e.g., social support, social comparison) through which Instagram usage directly impacts well-being.

Awards were granted of up to \$50k USD per awardee to fund projects of up to one year in duration. Current projects in receipt of Instagram Wellbeing Research Grants include *Impact of Instagram on wellbeing in adolescents vs young adults over time* (Dr Justine Gatt; UNSW and Neuroscience Research Australia) and *#InstaBIpositive: Supporting Positive Body Image on Instagram* (Dr Rachel F Rodgers; Northeastern University).

## **2.2 How we ensure diversity in what people see**

Social media is where people can turn to celebrate life's most joyful moments and seek support in some of the hardest. Online communities can provide invaluable support - they're a chance for people to bond over something they share. Facebook Groups and Instagram communities are a powerful tool in creating those communities and building meaningful interactions on Facebook and offline. With over a billion Groups on Facebook, everyone should be able to find a group that is meaningful to them. If they can't, we want them to start one and we're doing everything we can to help make that as easy as possible. Because meaningful groups can make a difference both on and offline.

Private groups can be important places for people to come together and share around a range of personal topics, like identifying as LGBTQI+ or discussing challenges around self-esteem and body image. But being in a private group doesn't mean that your actions should go unchecked. We have a responsibility to keep Facebook safe, which is why our Community Standards apply across Facebook, including in private groups. To enforce these policies, we use a combination of people and technology — content reviewers and proactive detection.

Facebook's Community Partnerships Team works with a range of communities focussed on safety, health and wellbeing. In the UK examples includes:

- Parenting Mental Health, a network of 20,000 parents helping one another as they support their children through mental health crises;
- Dope Black Dads, a network of 25,000 black parents committed to supporting each other and improving outcomes for black families.

### **2.3 Proposed/upcoming changes to boost diversity/limit harmful content**

At Instagram, we want the time our community spends on the platform to be positive, inspiring, and intentional. We know from our relationships and work with expert organisations that some people can feel a pressure to look or live a certain way because of social comparison to others. With this in mind we have worked with Internet Matters and Childnet in the UK to launch the ‘Pressure to be Perfect’ toolkit. The toolkit is about recognising that what you see posted by others is just one part of their life - that a single post or video rarely reflects all that is happening behind the scenes.

The toolkit is housed on the Internet Matters website and is available for download. There is one guide for young people and one designed to support parents.

<https://www.internetmatters.org/resources/helping-young-people-manage-their-onlineidentity/>

## **Regulation**

### **3.1 Facebook welcomes the aim of the Online Harms White Paper to ensure any new rules for the Internet preserve what’s best about it and the incredible benefits it has brought to the daily lives of billions of people, whilst protecting society from broader harms.**

Facebook has had rules about what content is and is not allowed on our platform for well over a decade. We develop these rules with a wide range of experts and partners and we fully understand the sometimes complex trade offs and decisions content moderation can involve. We have often said we’d welcome Government guidance, as well as the input of regulators – these are complex issues companies can’t tackle alone.

We want to work with the Government and Ofcom to develop the right regulatory framework that is effective and benefits everyone, and that achieves our common goal of making the online world safe. This needs to be a sophisticated framework that accounts for the different ways bad content can manifest online, the different nature of the platforms, and the need for flexibility to develop new solutions to detect and counter harms online. It also needs to align with the Government's aim for the UK to be a thriving digital economy.

We have long believed that empowering people to be digitally savvy is key. This is why we invest in a whole range of tools to give people control over their experience across Facebook’s platforms - everything from what kind of ads you see to managing your screen time - and we partner with a number of organisations to deliver digital literacy training, safety skills training and resources to young people, parents and teachers. The promise of a coordinated and strategic media literacy strategy from the Government is something we're excited to see.

There remain some areas in which we await further details on aspects proposed in the White Paper. For example, the White Paper proposes a “Statutory Duty of Care”. Companies like us



have immense responsibilities: we don't take them lightly and we fully accept that we should be accountable for enforcing standards on our platforms. Applying a legal 'duty of care' on to companies in relation to user content and speech is a first, and raises a number of important and difficult questions. Effective regulation should take into account the unique role that online platforms play and the difficulties involved in making judgements over what content to allow and what to remove.

A regulator who has access to information, expertise, clear accountability and transparency for their decisions could add value, but the powers of this regulator need to be clearly defined with proper safeguards. But there are some questions that still need answering – like what is the actual definition of 'harmful' content.

### **3.2 In many areas the combination of our Community Standards/Guidelines and our own advertising policies mean our platforms have stricter rules than the minimum requirements set out by the ASA**

In the UK, online advertising is subject to the strict rules set out in the Code of Non-Broadcast Advertising and Direct & Promotional Marketing (CAP Code), and administered by the independent body, the Advertising Standards Authority (ASA). Ads are regulated under a self-regulatory system, which means that the industry writes the rules that it is subject to. This sense of corporate social responsibility helps ensure that advertisers and platforms like Facebook have an interest in maintaining the integrity of the system.

The CAP Code not only requires that the content of advertising is “legal, decent, honest and truthful and consumer confidence is maintained” but also does not harm, mislead, or offend the public. Certain categories of products are subject to media placement restrictions, regardless of their content. This is to help ensure that children and young people are not targeted directly with age-restricted advertising, and that advertising is not allowed to appear in a medium where children or young people make up a large proportion of the audience.

### **3.4 Facebook believes that the ASA's ambition to work closely with online platforms to help protect people from irresponsible advertising, is the right approach**

We provide expert advice to the ASA/CAP on forthcoming issues relating to online advertising regulation, and there is a long history of effective partnership between the ASA and industry.

For example, the ASA recently ran a targeted ad on Facebook to raise awareness of its recent rulings on the prohibition against advertising Botox. The ad was seen by c. 1.4m people and viewed over 4.5m times (on average the ad was seen 3.37 times per-person). Over 2.5k people clicked the link, directing them to the ASA website to read the accompanying Enforcement Notice. As part of the project the ASA used its new monitoring technology to discover problem ads, and in turn report the ads to us to investigate and action their removal.

Where issues do occur and the ASA finds ads on Facebook that are in breach of the CAP Code it can report the offending material to us, through the Consumer Policy Channel (CPC), for action. The CPC is a reporting channel dedicated to effective engagement with consumer and advertising authorities who can send us take down requests directly to the channel. We work

with regulators around the world and receive take down requests about commercially-focused content, i.e. advertisements, commerce listings, and promotional content across the platform. If the content is found to be in breach of our policies or locally unlawful, we will take action against it. The ASA is onboarded to this channel and we understand from the ASA's perspective that this reporting flow has been effective in removing non-compliant material.

ASA/CAP can, with the help of platforms, raise its profile among hard-to-reach micro- and SME businesses. We can help run sector-specific awareness campaigns, as has been evidenced by the Botox ad campaign. We are also committed to working closely with the ASA and other regulators, who can bring breaches of the CAP Code to our attention.

June 2020