# Facebook—written evidence (FEO0095)

Facebook welcomes the opportunity to respond to the House of Lords Communications and Digital Committee's call for evidence into freedom of expression online.

We understand that the Committee is seeking to understand the current state of freedom of expression online, as well as considering how upcoming legislative and regulatory changes in the UK might impact rights to speech. As one of the most widely used social media platforms in the UK, we believe Facebook is well placed to comment on these questions.

In our submission below, we set our Facebook's approach to freedom of expression, how these principles relate to the policies that govern our platforms, and discuss how online regulation can balance free expression and what we believe will be the impact of the Government's proposals for upcoming regulation and legislation.

## 1.1 Introduction

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.

People use our products in multiple ways—over $1.8 billion was raised by our community to support the causes they care about in 2020 alone, 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 each day to help people express themselves and connect.

## 1.2 Voice and inclusion have gone hand in hand throughout history.

More people being able to share their perspectives has played an important role in building a more inclusive society throughout history. This is seen globally, where the ability to speak freely has been central in the fight for democracy worldwide. The most repressive societies have always restricted speech the most, and when people are finally able to speak, they often call for change.

Whether it's a peaceful protest in the streets, an op-ed in a newspaper or a post on social media, free expression is key to a thriving society. This is why, barring other factors, we lean toward free expression on Facebook. The internet has given people the power to share their stories directly and helped modern movements for change to reach many more people. Facebook is part of this trend, giving more than 3 billion people a greater opportunity to express themselves, exercise other rights like freedom of association and political participation, fulfil their right to livelihood and help others.

Movements like #BlackLivesMatter and #MeToo went viral on Facebook—the

hashtag #BlackLivesMatter was actually first used on Facebook—and it wouldn't have been possible for campaigners to organise on such a scale in the same way before platforms like Facebook existed. But while it is easy to focus on major social movements, we must remember that most progress happens in our everyday lives. It's the community groups organising support for isolated individuals during the pandemic (In April at the start of lockdown, more than 2 million people joined more than 2,000 COVID-19 support groups in the UK), or small businesses that now have access to the same sophisticated tools as big companies, so they can get their voice out and reach more customers, create jobs and become a hub in their local community.

People no longer have to rely on the traditional gatekeepers in politics or the media to make their voices heard. Just this year, young people in Thailand have been enabled by online tools to respectfully but bravely exercise their rights to protest and to criticise the monarchy. This power to express themselves at scale is a new kind of force in the world and one that has important consequences.

### 1.3 Free expression has never been absolute - some limits can and must exist.

Facebook is a platform for voices all around the world. The content we host is user-generated and moderated in accordance with our community standards when it is reported to us, or proactively detected as violating. Our rules are therefore developed to allow us to moderate many millions of pieces of content daily, across scores of languages, with robust guidance and training that seeks to minimize possibilities of discriminatory or arbitrary decision-making. However, some critical exceptions to free speech do exist.

The global standard defining freedom of expression (and with it, the right of access to information, and freedom of opinion) is defined in Article 19 of the International Covenant on Civil and Political Rights. This standard has developed from time-tested inputs of national governments, human rights experts, and civil society. It clearly indicates that freedom of expression should only be restricted to prevent imminent physical harm, to respect the rights and reputations of others, or for public health. Any restrictions should be: lawful, necessary and proportionate, meaning that they should be based on one of the above grounds; they should be necessary to protect rights in a democracy; and they should be crafted as narrowly as possible to meet their intended goal. Exceptions for public order are also permitted, but in states of emergency, restrictions on speech should reflect the guidance of the Siracusa Principles.[1]

In addition, Article 20 (2) of the ICCPR prohibits speech that incites hostility, violence or discrimination; incitement based on racial, religious, or national grounds is also prohibited. The more recent Rabat Principles[2] provide a multi-factor test to guide states in determining when such incitement must be criminally prosecuted. Article 20 (2) is interpreted and implemented in the legislation of multiple rights-respecting democracies. Facebook's Community

---

1 https://www.icj.org/siracusa-principles-on-the-limitation-and-derogation-provisions-in-the-international-covenant-on-civil-and-political-rights/

2 https://www.ohchr.org/en/issues/freedomopinion/articles19-20/pages/index.aspx

Standards on hate speech seek to implement this guidance. However, it is important to note that there is no global consensus on the borderline between offensive speech vs speech that should be limited under article (20)(2) of the Covenant - as with many other aspects of the online world, norms are evolving rapidly.

Facebook has sought to develop its Community Standards in alignment with these principles. We are assisted (and our performance is independently assessed) by our long-standing commitments as a member of the Global Network Initiative,[3] as well as by the UN Guiding Principles on Business and Human Rights[4] (UNGPs). The UNGPs likewise require businesses to undertake active human rights due diligence, and to prevent and mitigate adverse human rights impacts, depending on the strength of the nexus between the business and the harm. The Committee may wish to examine the UNGP framework in depth and will note its acknowledgement of the complexity in linking businesses to adverse human rights impacts.

Trying to put together a framework for speech that works for everyone, and effectively enforcing that framework across over 100 languages and cultural nuances, is challenging. Our Community Standards make clear that every policy we have is grounded in three core principles: giving people a voice, keeping people safe, and treating people equitably.

Frustrations do sometimes occur about our policies - both externally and internally - even when they are explicitly based on and repeatedly refer to internationally recognised human rights principles. But these are often the result of the inevitable tension between these three principles —as well as the fact that the meaning individuals derive from a single piece of content is highly variable, subjective, and context-specific. This is why we have welcomed regulation, in recognition of the fact that these are complex issues that we and other companies cannot tackle alone. Protecting rights, particularly freedom of expression, requires strong institutions across an entire ecosystem and we look forward to continuing to play our role in this.

## 2.1 We moderate content to protect freedom of expression not limit it

Facebook has had rules about what content is and is not allowed on our platform for well over a decade, which we call our Community Standards. We develop and iterate these rules with a wide range of experts and partners.

We do not believe that there must always be a tension between protecting the safety of our users and protecting their right to freely express themselves. In fact, we believe that moderating harmful content is an essential way of protecting freedom of speech. Doing so limits the malign influence of bad actors who look to undermine free expression with abuse and misinformation, and ultimately makes Facebook a more open and welcoming environment for our

[3] https://globalnetworkinitiative.org/
[4] https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

users.

Facebook therefore does not allow content that could physically or financially endanger people, that incites violence, intimidates people through hateful language, or that aims to profit by tricking people using Facebook. The way people use the internet is always changing, and so we also adapt our policies when necessary, making changes if it becomes clear that our existing policies don't go far enough, such as to defend against imminent violence or prevent physical harm provoked by false claims. This is in line with the internationally agreed limits on freedom of expression we discuss in section 1.3 and creates a safer environment for our users to connect and express themselves.

In writing our rules, we seek input from outside experts and organisations to ensure we understand the different perspectives that exist on free expression and safety, as well as the impacts of these policies on different communities globally. We make public the underlying principles we use to guide this stakeholder engagement on our website.[5] Every few weeks, our Content Policy team runs a company-wide meeting to discuss potential changes to our policies based on new research and data—and we've also invited academics and journalists to join these meetings to understand this process. We publish minutes[6] from these meetings, and we now include a change log so that people can track updates to our Community Standards over time.

We have over 35,000 people working in safety and security at Facebook with around half of these being content reviewers responsible for actioning reports made to us and enforcing our rules. To be transparent about the progress we are making against harmful content we issue a quarterly transparency report, which includes a Community Standards Enforcement Report[7] that shares metrics on how we are doing at preventing and taking action on content that goes against our Community Standards. These actions may include removing content or in other instances covering content with a warning screen.

In addition, we also regularly publish a report on actions taken in response to government requests for content restrictions or user data. We comply with Government requests for user data in accordance with applicable law and our terms of service and are committed to complete transparency in this area. We review every such request for legal sufficiency and may reject or require greater specificity on requests that appear overly broad or vague. This transparency[8] data is an essential accountability tool: the methodology has been reviewed by an independent academic advisory group and we are scrutinised by independent assessors[9] every two years on our progress in meeting our Global Network Initiative commitments to respect freedom of expression and the right to privacy as defined by the International Convention on Civil and Political Rights.

Our policies and Community Standards remain under constant review as we look

---

5 https://www.facebook.com/communitystandards/stakeholder_engagement

6 https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/#forum-minutes

7 https://transparency.facebook.com/community-standards-enforcement

8 https://transparency.facebook.com/

9 https://globalnetworkinitiative.org/2018-2019-company-assessments/

to create as safe an environment as possible for our users to freely share their views and build communities.

## 2.2 Content moderation is complex but the tools we use are improving constantly.

For much of Facebook's history we relied on manual removal of content that was reported to us by users. This approach enabled us to remove a lot of harmful content, but it meant we relied on those who were encountering bad content to share it with us. Moving from reactive to proactive detection of content at scale has only started to become possible in the last few years because of advances in artificial intelligence -- and because of the multi billion dollar annual investments we have made into this technology.

Today we use computers for what they're good at -- making basic judgements on large amounts of content quickly -- and we rely on people for making more complex and nuanced judgements that require deeper expertise and appreciation of context. 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. In the third quarter of 2020, our systems proactively identified 98.2% of the nudity we take down. While proactive detection of hate speech remains lower, we are making progress, and our proactive rate has climbed to 94.7%, from 68% at the start of 2019. We anticipate these figures improving further with advances in technology.

We recognise, however, that machine learning tools are less adept at handling context when detecting violating content. This is why using a combination of machine learning tools and human moderators is so important in ensuring our removal processes are not over-zealous and to the detriment of users' ability to freely express themselves. One of the primary functions of our human moderators is to help make our machine systems better in this way - minimising false positives is vital for ensuring we don't over-remove content, which could otherwise act as a limit on free expression. Every decision made by one of our human moderators is fed back to our classifiers to help make them more accurate.

It is important to note that even our machine learning tools can only begin proactive detection once content has been posted. Whilst we detect and moderate ever rising quantities before anyone ever sees it, there is always the possibility that someone will happen to be looking at the right place at the wrong time. Linguistic challenges are also difficult for our systems to assess. We both allow users to appeal content removals and conduct our own checks, to try to catch any mistakes that may have been made. Although we are committed to detecting as much violating content as possible before other users can see it, we are conscious that this figure can never be 100%.

Proactive enforcement does not change any of our policies around what content should stay up and what should come down. That is still determined by our Community Standards. Proactive enforcement simply helps us remove more

harmful content, faster.

### 2.3 It is inevitable that conflicts will sometimes arise between our commitment to free expression and to the safety or rights of our users.

It is important to note that the accuracy of a Facebook post is not in itself a reason to block it. For example, human rights laws extend the same right to expression to those who wish to claim that the world is flat as to those who state that it is round. It may be the case that false content also breaks other of our rules, but not always. In addition, freedom of opinion, so closely related to freedom of expression, "is the right to hold opinions without interference."[10] It is not derogable, and the authoritative guidance of the Human Rights Committee notes that all forms of opinion are protected. (General Comment No. 34,[11] para 9-10).

The internet is a new and unique technology that brings its own risks, including for free expression. Giving many more people a voice is dramatically empowering, but some people may use their voice for harm, such as organising violence or undermining elections by distributing manipulative misinformation. Ideas can also spread extremely quickly online - sometimes this is positive, but it can go the other way too. The internet is amazing in its ability to help people form communities that wouldn't have been possible before, but fears do exist that this has the potential to lead to polarisation.

Beyond these new properties of the internet, there are also shifting cultural sensitivities and diverging views on what people consider dangerous content. Furthermore, these are not always the same across the world. As a global platform, we have to attempt to take cultural differences and regional variations into account, while still maintaining our commitment to respect international human rights standards, and the associated authoritative guidance.[12] We do this through language and context specific interpretation and procedures that are consistent with, and implement, our global policies, such as lists of local slurs, and so forth. We complement our global rules with the insights from our extensive and growing network of Trusted Partners, NGOs worldwide who share their local expertise with our teams on difficult cases.

The challenge of developing community standards is further complicated by the speed with which social media has developed, and the resulting lack of coherent systems in national law. One example is the practice of bullying and harassment, and its (lack of) remedy in most national legal systems. Recent discussions in Canada[13] have started examining possibilities for remedy of bullying and harassment, however it is important to note that all sectors—business, government, legal, and education—have a great deal more to do to prevent or mitigate this harmful behaviour, one which so disproportionately impacts women and girls.[14]

---

10 https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf
11 https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf
12 https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf
13 https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/defamation-final-background-paper.pdf
14 https://undocs.org/en/A/HRC/38/47

## 2.4 Our policies are constantly evolving in light of multiple tensions as we seek to respect human rights standards while keeping people safe.

We adapt our policies when necessary, making changes if it becomes clear that our existing policies don't go far enough to defend against imminent violence or physical harm provoked by verifiable misinformation, or if they go too far.

For example, in January 2020, we applied this policy to COVID-19 to remove posts that made false claims about cures, treatments, the availability of essential services or the location or severity of an outbreak. We later began to remove claims that physical distancing did not help to prevent the spread of the virus and banned ads and commerce listings that implied that a product guaranteed a cure or prevented people from contracting the disease. In making this decision, we explicitly consulted the authoritative guidance of the UN Human Rights Committee on Freedom of Expression and the guidance of the Committee on Economic, Social and Cultural Rights (ESCR) on the content of the right to health and the role of access to authoritative health information as part of that right.

Following the news about the imminent COVID-19 vaccine rollout, we announced that we will remove false claims about vaccines where public health experts have told us these could lead to imminent real world harm. This could include false claims about the safety, efficacy, ingredients or side effects of the vaccines - for example, we will remove false claims that COVID-19 vaccines contain microchips, as well as conspiracy theories about COVID-19 vaccines that we know today are false: like specific populations are being used without their consent to test the vaccine's safety.

We focus on making sure that confirmed hoaxes don't go viral, with a particular attention on misinformation that could lead to imminent physical harm. More broadly though, we've found focusing on the authenticity and identity of the speaker works better than the content itself. This has driven our work on labelling state-sponsored actors, tackling fake accounts, and the rules we have to govern political advertising.

With regard to COVID-19, to tackle claims that don't directly result in physical harm, like conspiracy theories about the origin of the virus, we work with our network of independent fact-checking partners. Once a post is rated false by a fact-checker, we reduce its distribution so fewer people see it, and we show strong warning labels and notifications to people who still come across it, try to share it or already have.

In just the past year, we have grown our global network of fact checking partners to 80 organisations, working in 60 languages, and fighting misinformation for critical events like elections and COVID-19. To support the global fact-checking community's work on COVID-19, we partnered with the Independent Fact-Checking Network to launch a $1 million grant program[15] to increase their capacity. We know our efforts are working. From March to October

[15] https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking

of 2020, we labelled about 167 million pieces of COVID-19 related Facebook posts, resulting in a 95% drop-off in click-through to the underlying false content.

## 3.1 Facebook has long stated that the question is not whether to regulate but how.

As we've thought about the issues above, and how to balance these competing priorities, we have increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own.

Facebook established the Oversight Board to bring together global experts on these questions to provide binding decisions and guidance on how to enforce our rules. The Oversight Board is an independent body to safeguard its ability to make independent decisions and recommendations. More than a year was spent in global consultations to design the board's goals, structure, and charter, and guidance through a detailed independent human rights review.[16] The Oversight Board is financed through an independent trust and is functionally, legally, and financially independent from Facebook.

An initial cohort of board members,[17] diverse among multiple dimensions, was selected for their expertise and real world experience in freedom of expression, journalism, and other rights- related topics. These include former Danish Prime Minister Helle Thorning-Schmidt, and ex- Editor of the Guardian Alan Rusbridger. There is more information about the Oversight Board on its website, including relevant governing documents[18] and detailed information about the appeals process[19] (and other key issues).

On Thursday 28th of January, the Board published its first decisions on a selection of cases from around the world. We have immediately implemented their binding decision on the content, even where we disagree. We intend to review their decision and provide transparent updates over the next few weeks about the board's policy recommendations and how we will implement their decision across identical pieces of context with similar context where technically possible.

The board is the first and only institution of its kind, exercising the power to make binding decisions and codify exceptions to Facebook's policies where they see fit. We take responsibility for the consistent enforcement of our policies at scale, and we will continue to provide transparency around our processes and decisions.

16 https://www.bsr.org/reports/BSR_Facebook_Oversight_Board.pdf
17 https://oversightboard.com/meet-the-board/
18 https://oversightboard.com/
19 https://oversightboard.com/appeals-process/

### 3.2 We already seek guidance from existing regulatory regimes to guide us in this area.

Facebook, as a company, is expected to respect international human rights standards, as set out in the UN Guiding Principles on Business and Human Rights.[20] The expectation to respect differs from the duty of states, which is to protect human rights, as a result of longstanding treaty obligations and (in some elements) customary international law.

Facebook recognises the responsibility to conduct human rights due diligence and to respect the individual and human rights of the members of our diverse global community. We are dedicated to understanding the human rights impacts of our platform, and since 2018 we are the only large tech company to have publicly disclosed our human rights due diligence.

Examples of these steps include: our Myanmar, Sri Lanka, Cambodia, and Indonesia human right impact assessments; the human rights review of the Oversight Board; and the crafting of policies that allow us both to act against those who would use Facebook to enable harm, stifle expression, and undermine human rights, and to support those who seek to advance rights, promote peace, and build strong communities.

Indeed, it may be useful to note that the UNGPs focus on creating a systems approach to business respect for human rights. They specify the need to conduct due diligence; to involve rights holders; to prevent or mitigate adverse human rights impacts; and to offer remedies to those affected. They specify companies should prioritize harms based on scope, severity, and remediability, as well as offer an analysis of differing levels of responsibility (cause/contribute/linked through business activities). This framework is being developed further by initiatives such as the UN's B-Tech project.[21] Given the complexity, speed, and scale of social media company operations, we encourage regulators to consider the UNGPs and the value of designing a systems approach.

Facebook is also a proud member of the Global Network Initiative (GNI). The GNI works to advance the freedom of expression and privacy rights of Internet users worldwide. Companies that join the GNI agree to independent assessments of their record in implementing these principles and guidelines. In this biannual process, the GNI Board — which includes civil society organisations, academics, investors, and other companies — carefully reviews the findings of an independent, external assessment of each member company's practices, policies, and procedures, including human rights due diligence and employee training efforts. We are pleased that the GNI has certified that we comply with their Principles on Freedom of Expression and Privacy and we remain firm in our commitment to respect the rights of the people around the globe who use our services.

We also look for guidance in documents like Article 19 of the ICCPR and the

---

20 https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

21 https://www.ohchr.org/EN/Issues/Business/Pages/B-TechProject.aspx

authoritative guidance of the UN Human Rights Committee on freedom of expression, as discussed previously.

### 3.3 The UK's Online Harms proposals

Facebook welcomes the Government's proposals for addressing online harms and looks forward to working constructively with Parliament, the UK Government and Ofcom on the regulatory regime as it is brought into effect.

We have taken a number of steps both to learn from existing regulatory regimes when developing our policies and to introduce external oversight of them, as set out in previous sections. However, these are complex issues and we have long called for a more active role for governments and regulators to provide guidance on how these issues should be effectively addressed. By updating the rules for the internet, we can preserve what's best about it—the freedom for people to express themselves— whilst also protecting society from broader harms.

As with the approach of the UN GP framework, we agree that the most effective, balanced and scalable approach to tackling the challenge of online harms, is one that is focused on setting standards for systems and transparency and which is proportionate in accordance with the nature of the underlying harm(s) in scope. This regime will push services to build the right protections for their users while balancing the framework's guiding principles of protecting users' rights and supporting innovation.

An important part of the government's framework is the proposal to regulate posts and activity that could cause 'adverse physical or psychological harm' despite, in some cases, being legal.

This has never been done before, and represents a big change in the way people in the UK can use the internet. It will have substantial consequences that warrant careful consideration, and there is a risk that such a standard could introduce significant subjectivity in practice. The Government's decision to refer to the Law Commission the question of how to approach posts relating to suicide and self-harm demonstrates how difficult and nuanced these issues are. Ofcom will need to take a careful approach to enforcement in these areas to balance the different 'guiding principles' in the legislation. Facebook are keen to share our 15 years of experience in this space with the Government, Parliament, and the regulator where that may be helpful.

The Government's Response to the Online Harms White Paper consultation proposes strong safeguards to protect freedom of expression, to avoid companies being overly cautious and removing too much content. We look forward to seeing further detail about how these measures will need to be applied, particularly in light of the ever-evolving nature of technology as a tool to aid content moderation. In addition, we are keen to understand the expectations around the proposed requirement for platforms to not 'arbitrarily' remove opinion. As stated above, we always endeavour to enforce our policies consistently and transparently, and welcome clear guidance as to how companies can continue to develop and enforce their own standards in this regard.

We hope to see further details to explain the above points when the Draft Bill is published soon. We look forward to working with Parliament and Government to make the Online Safety Bill as effective as it can be.

### 3.4 Other measures in the Online Harms proposals, beyond content moderation, may have a bearing on the status of free expression online.

We have long believed that empowering people to be digitally savvy is key, therefore the promise of a coordinated and strategic media literacy strategy from the Government is something we're excited to see, and we were glad to see continued commitments to online media literacy in the Government's Response to the Online Harms White Paper consultation.

We have invested in a range of tools and partnerships to help young people think critically and share thoughtfully online. In August 2018, we launched our Digital Literacy Library,[22] a resource for educators looking to address digital literacy and help these young people build the skills they need to safely enjoy digital technology.

We have supported a range of work to help young people in the UK to spot false news and improve their digital literacy. In 2020 we launched our partnership with the Economist Educational Foundation's Burnet News Club. This programme provides teachers in non- selective state schools across the UK with resources to run weekly news clubs. These news clubs enable pupils to develop their news literacy, critical thinking and communication skills by encouraging well-informed, open-minded discussions about current affairs. Our support enables schools to access the programme at a subsidised rate, enabling the Foundation to increase the number of schools they can work with by 50%.

I hope that this submission has helped to provide further clarity on the way Facebook approaches the important task of keeping our users safe while protecting their rights. This issue will be the subject of considerable scrutiny and debate in Parliament and in society at large as the Online Safety Bill passed through Parliament this year. We look forward to contributing to these debates, including those with your committee, in the coming months.

*29 January 2021*

22 http://facebook.com/safety/educators