FACEBOOK

Facebook's Submission to the DCMS Sub-committee on Online Harms and Disinformation Committee's inquiry into Online safety and online harms — September 2021

Introduction

Regulating the content people generate online, through videos, posts, comments, Pages, and Groups, requires new and innovative regulatory frameworks. These frameworks must ensure the online world is a safe place but also one where people's fundamental rights to privacy and expression are respected. A place where online service providers act reasonably and proportionately, taking their share of the responsibility to ensure this complex balance is struck.

Facebook will continue to be a constructive partner to governments as they weigh the most effective, democratic, and workable approaches to meeting this challenge. Over the last four years, Facebook has supported the UK Government's development of the Online Harms framework through written submissions, ministerial discussions, and multistakeholder in-person technical sessions.

We share the UK Government's stated policy objectives, to make the internet safer while protecting the vast social and economic benefits it brings to billions of people each day. It is important not to lose sight of those benefits, but to have them squarely at the heart of how the UK approaches the Online Safety Bill. Global connectivity has improved economies, grown businesses, reunited families, raised money for charity, and helped bring about political change.

Millions of UK small businesses use Facebook to reach customers and grow, and almost 1 in 4 of these small businesses say that the growth they have achieved using Facebook, its free tools, and the personalised advertising that it enables, has led them to hire at least one new employee. 35 million Brits now use Facebook Groups to connect every month, and 39% of Brits say the most important group they are a part of now operates primarily online. During the pandemic, over one and a half million people in the UK joined Coronavirus Mutual Aid groups on Facebook, and globally over two billion people have visited our COVID information centre, which features the latest updates from health authorities including the NHS.

But at the same time, the internet has made it easier to create, find and spread content that could contribute to harm, like hate speech, misinformation, and terror propaganda. At Facebook we have sixteen years' experience in tackling these issues through establishing policies, building tools and technologies, and producing guides and resources, all in partnership with experts both within and outside our company. We want to share what we have learnt about what works, and as the Committee has asked, to share our experience of offering our services under other, similar laws in other countries

If designed well, new frameworks for regulating harmful content can contribute to the internet's continued success by articulating clear, predictable, and balanced ways for government, companies, and civil society to share responsibilities and work together. Designed poorly, these efforts may stifle expression, slow innovation, quickly become redundant and create the wrong incentives for platforms.

FACEBOOK

The UK has a valuable opportunity to lead the thinking on internet regulation and write new laws which align its regulatory environment with its digital ambitions. With the right reforms this will help unlock the potential of its world-leading tech scene, which Facebook via our significant presence in London has been at the heart of for more than a decade.

Through detailed Parliamentary scrutiny - including that of the DCMS Select Committee - the UK Government's Draft Bill stands a real chance of meeting these ambitions.

Below we set out our views on the principles of the Bill framed around its **overall objectives, tensions, and workability** and refer to the questions asked by the Committee within these sections**. Our recommendations to the Committee are presented in Bold**, and collated at the very end of this submission.

FACEBOOK

4. International consistency	*Alignment*

1. Aspects we welcome

Facebook welcomes the Draft Bill's attempts to establish a systems-focused framework, requiring service providers to use proportionate systems and processes to address harmful content. We believe this approach is the best way to ensure an appropriate balancing of safety, privacy, freedom of expression, and other values, and have long been on record calling publicly for legislation following this model.[1]

In this section we highlight several aspects of the Draft Bill which we believe contribute to building a robust and effective regulatory framework that can command public trust: the appointment of Ofcom; transparency and audit, researchers' access to information, user agency, and age assurance.

Ofcom as the regulator

Firstly, the Government's decision to appoint Ofcom as the regulator will bring many years of knowledge and expertise to bear on the enforcement challenges that the Bill presents. Facebook has worked constructively with Ofcom in the past on media literacy initiatives, and more recently as the organisation has begun to prepare for its future role. We are confident that Ofcom will approach this new role with a high degree of diligence and professionalism, strengthening the effectiveness of the regime and public confidence in it.

Transparency and Audit

We also welcome the Draft Bill's focus on increasing transparency through greater reporting obligations. We believe this can play a vital role in helping to ensure there are common guidelines and benchmarks to develop a level of consistency and predictability among services and foster a climate of accountability.

In Facebook's case, bringing greater transparency to our policies, tools and enforcement has been a priority of ours for many years. We are committed to sharing meaningful data so we can be held accountable for our progress, even if the data shows areas where we need to do better.

Each quarter we publish our Community Standards Enforcement Report. This report provides metrics on how we enforced our policies over that period. The report now covers 13 policy areas on Facebook and 11 on Instagram. A major focus of the development of this report has been identifying and

[1] 'Four Ideas to Regulate the Internet', Facebook, 2019, https://about.fb.com/news/2019/03/four-ideas-regulate-internet/

producing metrics that provide the most meaningful insights in terms of measuring both forms of harm and the effectiveness of our policies against them.

The amount of hate speech content we remove each quarter has increased over 15 times on Facebook and Instagram since we first began reporting this, but more tellingly, the prevalence of hate speech on Facebook has decreased for three quarters in a row since we first began reporting it - i.e., how many times a given type of harmful content is likely to be viewed as a proportion of the overall views on that service. This is due to improvements in proactively detecting hate speech and ranking changes in News Feed, which has decreased the amount of hate speech users encounter.

We believe, Ofcom can play a pivotal role in ensuring adequate transparency across the whole industry, by reviewing annual reports and requesting certain information from companies about how their systems and policies operate.

Report on researchers' access to information

Clause 101 of the Bill provides for Ofcom to prepare a report about how researchers into online safety are able to access information from service providers. Ofcom will have to consult with service providers, the research community, and the Information Commissioner in preparing this report, and later may make guidance to support the carrying out of this research.

Facebook has a long history of seeking to make privacy-protected data available to support research. Our Data for Good programme sits at the heart of this work and aims to provide timely insights and relevant datasets to help improve how non-profits do their work, how researchers learn, and how policies are developed. When data is shared responsibly with the communities that need it, it can improve wellbeing and save lives.

Age Assurance

We welcome the special focus within the Draft Online Safety Bill on keeping children and young people safe online. Facebook believes it is important to strike the right balance between giving young people the ability to exercise their digital rights through our services while also keeping them safe. Where necessary and appropriate, we have implemented further safeguards for young people, striking a balance between protecting young people and facilitating their connection and development in the digital environment.

To achieve this balance, we recognise the role of proportionate and risk-based age assurance, among other safety and privacy safeguards, in helping to ensure that young people receive an age-appropriate experience. However, it is also important to recognise that age management is a complex challenge for industry, regulators, parents, and children, requiring thoughtful solutions that must protect young

people's privacy, safety, and autonomy, without unduly restricting their ability to access information, express themselves, and build community online.

We support closer industry collaboration to develop effective measures to ensure young people consistently receive age-appropriate experiences across the online ecosystem. In tandem, we welcome the opportunity to work collaboratively with the government, regulators, experts, industry, parents/caregivers, and young people to agree common age verification standards that can ensure consistency and trust in age assurance solutions.

2. Overall Objectives

Q How has the shifting focus between 'online harms' and 'online safety' influenced the development of the new regime and draft Bill?

Systems Focused Approach

As outlined above, Facebook welcomes and shares the Government's overall objectives for the Bill, which are to make the UK the safest place to be online, while supporting the use of technology to build an inclusive, competitive, and innovative future digital economy in line with the Digital Strategy. While the title of the Bill has shifted from 'online harms' to 'online safety', it still attempts primarily to establish a systems-focused framework, requiring service providers to use proportionate systems and processes to address harmful content, which we support. We believe this approach is the best way to ensure an appropriate balancing of safety, privacy, freedom of expression, and other values, and have long been on record calling publicly for legislation following this mode.

However, as currently drafted, we believe the proposed legislation risks failing to deliver the Government's policy aims effectively. The main reasons for this are the introduction since the consultation on the Online Harms White Paper of significant additional complexity and competing priorities. The best way to realise the Government's ambitions is to simplify and focus the proposed regime, and this note proposes a number of practical steps to achieve this.

Since the publication of the Internet Safety Strategy Green Paper in 2017,[2] the Government has been consulting industry, civil society and others on the development of a framework for tackling online harms. That framework would hold services to account for enforcing their own rules. It would demand online services do this by having effective systems and processes in place to prevent harm to their users. In order to ensure these proposals were as future proof as possible, and as broadly applicable as

[2] 'Internet Safety Strategy Green Paper,' HM Government, 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf

possible, the Government's approach would be underpinned by some key principles, including that companies should take risk-based, reasonable and proportionate steps to address online harms.

As the then Secretary of State Jeremy Wright said at the launch of the Online Harms White Paper in April 2019, "*I believe the United Kingdom can and should lead the world on this. Because the world knows we believe in innovation and we believe in the rule of law too. We are well placed to act first and to develop a system of regulation the world will want to emulate*".

Facebook's view is that the UK does have the opportunity that the then Secretary of State described, and that the model proposed at that time is the most effective approach to addressing online harms. Put in its simplest terms, the intention was to ensure that all online services within scope analysed the risks which could appear on their services and then took reasonable steps to mitigate them through their systemic approaches to managing content.

This model would avoid the mistakes seen in some other jurisdictions where policymakers try to prescribe what must be done in relation to certain specific harms or focus down on individual instances of harmful content online rather than on systems and the overall approach. As Ofcom Chief Executive Melanie Dawes said in December 2020 "*We won't be responsible for regulating or moderating individual pieces of online content. The Government's intention is that online platforms should have appropriate systems and processes in place to protect users; and that Ofcom should take action against them if they fall short. We'll focus particular attention on tackling the most serious harms, including illegal content and harms affecting children*."

The Draft Online Safety Bill (OSB) as published in May 2021 remains, in many respects, an attempt to establish such a systems-focused framework, placing on service providers "*a duty to operate a service using proportionate systems and processes designed to minimise…"* the presence, duration, and dissemination of harmful content.[3] This duty is to be based on related risk assessments.

However, elements of the draft legislation have drifted away from this original intention, and we believe risk introducing complexity and incoherence. For example, the draft Bill departs from the original principles by now requiring online services to find, identify and then act on (including to protect) specific content which is either from a journalist or citizen journalist; to protect posts on areas of political controversy; and to not discriminate against a particular political viewpoint. It is unclear how services can do this without having to make case by case decisions on different individual pieces of content—which will put either service providers or Ofcom in a position which the original Government proposals rightly sought to avoid.

The current proposals create fundamental ambiguity about what successful compliance would look like in a regime which moves between a systems focus and a case by case focus, due to the nature of the

[3] 'Draft Online Safety Bill', HM Government, 2021, cl.9(3), cl.10(3), cl.11(2)

obligations associated with different categories of illegal and harmful content. This ultimately threatens to undermine the Bill's good intentions. In the next section we comment on the fact that it is currently unclear how the multiple duties in the draft Bill interact, and how the inevitable tensions and trade-offs are to be balanced—for example, when a democratically important piece of content might also be harmful.

Facebook's experience is that the more prescriptive the regulatory requirements and the greater the focus on individual pieces of content, the slower and more complex the decision making for content reviews. Ultimately, under the risk of large penalties, an incentive is created for the mass removal of *any* content which *may* fall foul of these rules. This runs counter to the Online Safety Bill's ambition to protect people's rights to freedom of expression.

These risks in the draft Bill are further compounded by the definition of harm being so specifically tied to individuals. It is not clear how Ofcom will be able to use this definition to measure the effectiveness of systems and processes of a service, while at the same time remaining above involvement in case by case individual content decisions.

Narrow requirements focused on specific content could also quickly become outdated and restrictive in the rapidly changing environment of online communication, raising the question of how future-proof the framework may be. Rather than using the Bill to set out an overall framework, we believe too many complex and ambiguous exemptions and specifics have been put on the face of the Bill, some without consultation, raising the risk that the framework could prove difficult to implement effectively. Detailed provisions like this would be more appropriately developed and maintained by the regulator, Ofcom.

Recommendation: The DCMS Committee should encourage the Government to ensure that the final Bill aligns fully with the systems- and process-based framework which it has developed with stakeholders since 2017. The proposed definition of harm and certain provisions in the Draft Bill as written risk derailing the whole approach towards a focus on individual pieces of content and should be revised to ensure overall coherence and effectiveness. Further details on implementing the framework should be left to Ofcom.

Q Does the draft Bill focus enough on the ways tech companies could be encouraged to consider safety and/or the risk of harm in platform design and the systems and processes that they put in place?

Platform Design and Media Literacy

It's important that laws regulating harmful content take account of the fact that speech norms vary both across countries and between communities in a country and give sufficient consideration to the role of platform design and user agency in promoting online safety. In light of this, Facebook welcomes

the Draft Bill's recognition that efforts to address potential harms should consider the question of "*how the design and operation of the service (including the business model, governance and other systems and processes) may reduce or increase the risks identified*."[4]

Ensuring that people have the ability to control their own experience online is at the heart of what we do at Facebook. This means giving them both the ability to personalise their social media feeds, and also to take steps to ensure their own safety. People who use Facebook and Instagram can tailor the content that appears in their News Feeds in multiple ways, from easily choosing to snooze, unfollow, or block unwanted accounts to showing their News Feeds indifferent orders on Facebook, including chronologically. They can control the ads and other content they are served and amend or remove the information we are using to tailor their experience.

In terms of safety, our Community Standards set out clearly the types of posts and content we will remove from our services. However, beyond a certain point users play a critical role in ensuring online safety. Put simply, while there are a wide range of harms which can and should be addressed by companies, there are others where it would be inappropriate to centrally enforce standards, and where the user experience would be best left to individuals to control. Profanity is a good example. Some people will find swearing very offensive and even harmful, some do not. Blanket removal of swearing would lead to an undesirable level of censorship of legitimate speech. Therefore, it is the better approach to give users the option to filter out that type of content where it is their preference, rather than have our central systems do it.

The other benefit of user controls is that they can prevent harms from happening in the first place, because an individual knows what is likely to be offensive or harmful to them personally, or who is likely to offend them. On Instagram for example, people can use filters to prevent others from leaving potentially offensive comments that use words, phrases, or emojis they don't want to see. This year, we expanded this feature to also cover Direct Message requests, so that messages containing certain words, phrases or emojis will be automatically filtered to a separate hidden requests folder. These tools work alongside existing features such as: the ability to restrict and block individuals (including blocking future accounts a person may set up); the ability to manage multiple unwanted comments in one go—whether that's bulk deleting them, or bulk blocking the accounts that posted them; limiting who can send you Direct Messages; and automatically hiding comments and DM requests from people who don't follow you, or who only recently followed you.

In-app tools are also important for changing behaviour, to again prevent harms from happening in the first place. For example, we saw a meaningful decrease in offensive comments after we started using AI to warn people when they're about to post something that might be hurtful. After testing we updated this tool and found that in one week we showed warnings about a million times per day on average to people when they were making comments that were potentially offensive, and of these about 50% of

[4] Draft Online Safety Bill, HM Government, 2021, cl.7(8)(g), cl.7(9)(g), cl.7(10)(g), cl.9(10)(g), cl.19(3)(d), cl.19(4)(d).

the time the comment was edited or deleted by the user. The example illustrates that, at times, legal but potentially harmful content can serve as a moment of education and behaviour change which could not be achieved by incentives that only require the wholesale removal of harmful content. The same can be said of applying warning screens or fact checking labels—these measures give users agency, control and more insight into what they are seeing and how to help ensure their own safety as well as improve the safety of others.

It is welcome that the Bill contains a clause relating to media literacy, which gives Ofcom a duty to promote media literacy through educational activities and guidance (with powers for the Secretary of State to give directions on this guidance). In Facebook's experience, these activities are an essential component to any framework seeking to proactively address harm online, and we look forward to seeing Ofcom's work set out in more detail, including how services might work with Ofcom in meeting their respective duties.

Facebook has worked closely with the UK Government in the preparation of its 2021 Online Media Literacy Strategy, and we will continue to work with the Government and our partners to support programmes that educate and empower internet users across the UK to manage their online safety.

Recommendation: The DCMS Committee should continue to champion the importance of user agency and media literacy as integral parts of the UK's online safety landscape, and should encourage Ofcom and DCMS to move ahead with implementing the Online Media Literacy Strategy while the Bill is passing through Parliament.

3. Tensions, contradictions and workability

Q Are there any contested inclusions, tensions or contradictions in the draft Bill that need to be more carefully considered before the final Bill is put to Parliament?

Tensions and Contradictions

There are tensions and contradictions in what the Bill asks services to do, without sufficient clarity on how these should be balanced. The Bill's core safety duties require services to take steps to find and remove illegal and harmful content. However, these duties have been placed into tension with a series of countervailing obligations to take care *not* to remove content where it falls under certain categories. This includes content 'of democratic importance', or content 'created for the purpose of journalism', or where to do so would conflict with rights to freedom of expression or privacy—but crucially, these terms are poorly specified, which risks the balance tipping in favour of over-enforcement.

In operating our services, Facebook is well used to balancing competing values, such as those between voice, authenticity, and safety, privacy, and dignity, by drafting and refining our Community Standards and building tools to enforce them. We have set out our approach to reaching clarity and consistency in striking this balance publicly in our Newsroom.[1] However, the Government has not yet provided a similar model for how the different duties in the Draft Bill should be interpreted or reconciled, which raises two key points.

Firstly, due to the sheer volume and diversity of content decisions that services like Facebook must make every day, we have to use at-scale techniques to make many of those decisions. To make this possible, our rules must be extremely clear, rigorous and consistent in how they balance the competing values that underpin our Community Standards. The less clarity there is in the framework about how different duties should be balanced, the greater the likelihood that one of two things can happen: harmful content stays up too long as services need to use slower processes to work out which of the competing considerations to prioritise; or all of it is removed to mitigate risks of enforcement action, and legitimate speech is thereby censored.

The second issue is that without directions for how the duties should be interpreted or reconciled, it is not possible to predict how Ofcom will hold services to account for their efforts. According to what principles will Ofcom judge the attempts by services to balance these competing requirements? Without this clarity, neither Parliament nor the services affected can properly scrutinise the framework, estimate its effectiveness, or prepare to implement it.

Recommendation: The DCMS Committee should call on the Government to fully define the conflicting duties placed on services in the Bill, and provide further clarity on how services will be expected to balance them, either through draft Secondary Legislation or by setting out how Ofcom will evaluate services for striking these balances.

Workability

In this section, we highlight specific areas of the Bill that we believe merit amendment before the Bill is introduced to ensure workability.

Content in Scope
Democratic and Journalistic content

The draft Bill states that Category 1 services must have systems and processes in place that ensure the importance of democratic free expression, diversity of political opinion and free and open journalism are taken into account when making decisions about taking down or restricting access to such content. We are concerned that the Government is putting obligations on private companies to make complex and real time assessments about what constitutes journalistic and democratic content which could be

impossible to implement consistently. Private companies should not be the arbiters of what constitutes journalism or what is democratically important.

Democratic content

In the draft legislation, content of democratic importance is defined extremely broadly, as content that "*is or appears to be specifically intended to contribute to democratic political debate*". 'Democratic political debate' could cover a very wide range of topics, and the Bill does not explain when content should be considered to be contributing to political debate (or when comments and posts on a political issue might be anything other than democratic).

In addition, because the test is whether the content is (or appears to be) 'specifically intended' to contribute – not actually to contribute – it is unclear whether any content could be excluded from the protections being offered by this part of the Bill without knowledge of the intention of the person posting it. In many cases, it will be impossible for a company to be clear whether the user is specifically intending to contribute to a debate, and leaving these millions of subjective case-by-case decisions to in-scope services to decide is unworkable in practice.

Further, it is unclear where and how the limits of 'a diversity of political opinion' are to be drawn, and whether political opinions that are extremist but not unlawful (e.g. fascist statements by a far right organisation) must be treated in the same way as mainstream views. This is likely to be a particular problem where content which meets the threshold for 'diverse political opinion' or 'democratically important content' is also potentially harmful to adults or children (e.g. because they are extreme or divisive statements), resulting in tensions between this duty and the duties to protect user safety. The Bill offers no clarity about the process a company should follow in this scenario.

Journalistic content

In a similar manner, the duty to protect journalistic content is vague and risks introducing a significant element of subjectivity and therefore complexity. The draft bill defines 'journalistic content' as content that 'generated for the purposes of journalism', and which is 'UK-linked'. This journalistic content could be either regulated content (posted by users) or 'news publisher content'—a special category created with the intention of preventing the legislation from incentivising censorship, but the Government's accompanying press release suggests that citizen journalists' content will have the same protections as professional journalists' content.

These complex, circuitous, and potentially overlapping definitions will make it extremely challenging for platforms to design systems and processes that rigorously and consistently identify and appropriately handle the different forms of content. This is especially the case when it comes to content from users who assert that they are operating as citizen journalists. We have already seen instances where some users claim they are citizen journalists in an attempt to reduce the likelihood of Facebook (and other

platforms) taking action against them or their content. It will also be placed in direct tension with the other duties such as addressing legal content that is harmful to children or adults, as we set out in section 1 above. In-scope services must disclose in the Terms of Service the methods they use to identify journalistic content, which will make the system easier to game and is likely to lead to an increase in content from supposed citizen journalists as people try to ensure their content is given protections.

Furthermore, the dedicated and expedited complaints procedure set out in Clause 14(3) is also likely to lead to instances where users argue they generated content for "*the purposes of journalism*" in order to gain access to a faster complaints procedure. This is another instance of the way this provision conflicts with the objective of designing a systems-based approach.

In summary, the risks that arise from these provisions as drafted are twofold. First, the terms used and the basis on which service providers must judge these complex decisions are vague and open to subjective interpretation. Without clear definitions and guidelines the application of these requirements could become arbitrary and inconsistent across platforms, and will ultimately act in opposition to the Bill's objectives. Secondly, these requirements may inadvertently give a level of protection to bad actors or harmful content, which may expose users to an increased risk of harm.

Recommendation: The provisions relating to journalistic and democratic content should be removed entirely from the Bill, as they add additional risks to the framework rather than reducing them. If they are retained, the Government must set out clearly defined parameters around the definitions of all content types, and the connected duties, so that in-scope services have much greater clarity on their responsibilities.

Fraud

The Draft Bill does not contain any provisions on tackling online fraud. However, in the press release announcing its publication, the Government stated that fraud from user-generated posts would be included in the final text. Facebook supports the Government's ambition to make the UK the safest place to be online, including by addressing the potential harm to users from online fraud. However, we are of the view that the Online Safety Bill is not the best mechanism to address this.

Online fraud is complex, it can manifest in many ways—through content posted by an individual user or a company, through organic (user-generated) content or paid content (advertisements). Potentially fraudulent online content can look benign, and the user journey can involve many touch-points both online and offline before the fraud takes place. As a late addition to the Bill, there has been little detail given about the definition of fraud, the type of fraudulent activity that will be subject to the Bill and whether the definition of 'user-generated' includes users who are companies. Ofcom will need to grapple with these complex issues if fraud remains in scope of the Bill. Furthermore, the Bill was not written with the intention of regulating economic crimes, it was designed to focus on user-generated

content causing 'physical or psychological harm'; it is hard to see how this definition can be expanded without making the framework unworkable.

Regardless, the online harms regulator Ofcom cannot tackle the issue of online fraud alone—it will require multi-stakeholder collaboration across Government, the financial services industry, the tech industry, and law enforcement, and both legislative and non-legislative approaches. Expert stakeholders are already looking to address online fraud through the Home Office's Fraud Action Plan, DCMS's commitment to use its upcoming Online Advertising Programme to specifically look at paid-for advertising as a vector for fraud, and through collective initiatives like the Online Fraud Steering Group. The OFSG has been established as a cross-sector group to reduce the threat of online fraud over a minimum period of six months, and charged with forming targeted responses and specialist solutions. These established channels align better with the existing regulations and regulatory bodies that tackle fraud offline, and can provide ready-made and potentially quicker solutions than the Bill.

Recommendation: Instead of tasking Ofcom to address online fraud through the Online Safety Bill, the Government should involve Ofcom in the ongoing cross-Government work to tackle fraud, including the Fraud Action Plan and the Online Advertising Programme. Ofcom and the financial regulators should work together to ensure any proposed steps to tackling fraud through user-generated content are workable and align with Ofcom's OSB obligations.

Services in Scope

Categorisation of Services

The Bill sets out a 'differentiated' approach to in-scope services, placing additional duties on firms that are designated as 'Category 1', so that only these companies have to address 'legal but harmful' content.

It is right that regulation takes account of the varying risks and capabilities of different services. However, the Bill makes a flawed assumption by using size as a proxy for harm. The Government uses the term *"high-reach, high-risk"* when describing Category 1 firms, implying that services with the largest user base are inherently higher-risk for lawful but harmful content.[5] The Bill itself also tasks Ofcom to focus on the 'size and capacity' of the service as of equal importance to the service's own risk assessment when determining the proportionality of various safety steps it takes (cl.10(6)a).

We believe this is mistaken. Harm travels between services. Frequently, hateful, dangerous, or false narratives emerge on our platforms having first been developed and spread on smaller or less

[5] Memorandum from the Department for Digital, Culture, Media and Sport and the Home Office to the Delegated Powers and Regulatory Reform Committee, HM Government, 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985030/Delegated_Powers_Memorandum_Web_Accessible.pdf

sophisticated services. These services are less stringently moderated and have less developed technology to prevent the proliferation of harmful content than Facebook and other larger services. We share the concerns raised by several anti-hate speech organisations that the Online Safety framework must not focus on legal harms only as they appear on the largest platforms, at the expense of other parts of the internet where evidence suggests these harms flourish.

Setting up the framework in this way could create incentives for bad actors to move to smaller, less regulated platforms, hence masking the problem and putting harm out of reach of many of the Bill's provisions. The Government should consider a more equitable, risk-based approach that addresses risks where they appear instead of assuming that the largest services are the main problem.

Recommendation: The Bill should set a more sophisticated method of defining 'Category 1' services, that look not just at size and functionality, but at presence of harms and mitigating actions taken by the service. Ofcom should be given additional powers to carry out research into the prevalence of legal harms on small services, and the power to bring services into the scope of Category 1 for a period even if they do not meet the necessary thresholds. Details should be provided much earlier about any thresholds to give providers of different services clarity sooner.

Private Messaging

We welcome that the regulatory framework contains certain general provisions relating to privacy. There is the overall requirement to consider user privacy, and services must show what steps they have taken to protect privacy when implementing safety policies. However, given the other significant and novel elements of the Bill, especially those which seek to regulate legal speech, these are fairly weak references to the right to privacy and the due consideration companies should give it. As such, we believe the Bill could be improved by significant strengthening of this language to ensure it is given equal weight as the other aspects of the framework.

Recognising that privacy is a fundamental right, the 2019 Online Harms White Paper made clear that the Government understood public and private communications are different. Therefore, harms in private communication should be subject to a tailored set of requirements to ensure steps to address harms were effective and proportionate. In our response to the consultation on the White Paper, Facebook supported this approach and expressed the view that any definition of 'private communications' should distinguish between services which host and serve *public* user generated content, and *private* messaging services. Making this a very clear distinction seemed to be the minimum response required given that the Government itself noted, in its summary of written responses to the White Paper consultation, that *"overall respondents opposed the inclusion of private communication services in scope of regulation*".[6]

[6] Our emphasis, 'Online Harms White Paper—Initial Consultation Response', HM Government 2020, https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response#chapter-one-detailed-findings-from-the-consultation

The draft Bill appears not to have reflected that consultation feedback, and has also not maintained the distinction between public and private services. The single definition of 'user-to-user' services now encompasses all types of internet service through which users can generate, upload, share, or encounter content generated by other users. The only situation in which the draft Bill distinguishes between private and public parts of a service is in Ofcom's ability to require 'use of technology' notices to address certain types of content, which we discuss in greater detail below.

We believe that the Government's change of direction to treat public and private communications identically is the wrong approach. Private messaging (such as SMS, messaging or email) is fundamentally different from public social media in many crucial ways.

Purely private online messaging services, such as WhatsApp, are designed to allow people to stay in touch with friends, family and co-workers, privately and securely, anytime and anywhere. Most online private messaging services have similar features to WhatsApp including:

- You must have someone's phone number to contact them
- You cannot search for people or content in the way that social media platforms allow.
- There is no obligation to upload a profile photo, name or other personal information, and you control who can see this information if you do.
- Private messaging does not suggest or promote any content outside those to whom it was sent, or suggest 'people you may know'

Privacy and data minimisation are at the heart of most messaging services, especially encrypted services. WhatsApp for example deliberately collects only limited categories of personal data from users. Your personal calls and messages are secured, and only the sender and receiver can access the contents of messages.

Lastly, conversations in private messaging carry a high expectation of privacy. On WhatsApp, 90% of chats are between two people and the average group size is fewer than ten. Actions intended to prevent harm in public spaces would be entirely inappropriate and in contradiction with existing rights in private messaging.

In the Final Response of December 2020, the Government recognised that the framework would apply to public communication channels "*and services where users expect a greater degree of privacy*", and in recognition of this, stated that Ofcom would set out how companies can fulfil their duty of care in codes of practice, *"including what measures are likely to be appropriate in the context of private communications"*. We believe that returning to this approach would enable a more appropriate balance between users' rights to privacy and society's expectations of safety online.

Recommendations: If no distinction is to be made between public and private online services, then private communications should be removed from the scope of the Bill entirely (as was the view of the

majority of respondents to the original consultation). Alternatively, the Government should revert to its original approach and set out differentiated obligations on services based on their privacy or accessibility, and taking account of relevant technical capabilities.

Use of Technology notices

We are concerned by the powers set out in clause 64 of the Draft Bill, which gives Ofcom the power to issue a 'use of technology' notice to a regulated service requiring that service to use a specified technology to identify and 'take down' terrorism content or child sexual exploitation and abuse (CSEA) content. In relation to terrorism content, this power is limited only to 'public terrorism content', whereas for CSEA content the power extends to 'content present on any part of the service (public or private)'.

There are a number of problems that arise from this power, in addition to the fact that they move the framework away from the 'systems and processes' approach that is meant to be at the heart of the Bill, as we have set out above.

Firstly, the reference to 'taking down' content from private services does not make sense in the case of private communications. Unlike in a public social media service, no content has been 'put up' when a message is exchanged between users in private. The only feasible interpretation of this wording is that, if a 'Use of Technology' notice was issued, all private messages would need to be monitored in some way and anything which breaches the rules should be removed. This position is impossible to reconcile with the privacy obligations elsewhere in the Bill, and is inconsistent with the hosting provider defence under the e-Commerce Directive Regulations 2002. The Online Harms White Paper stated that the regulations would be compatible with the e-Commerce Directive, and the Full Government Response repeated that "*companies' liability for specific pieces of content will remain unchanged*"[7].

Second, these steps would significantly undermine individual privacy and be incompatible with end-to-end encrypted services. Encryption currently gives the highest degree of privacy and data security to billions of people. The draft Bill does not clarify exactly what a 'use of technology' notice could seek to compel, but it is reasonable to infer that it would involve gaining access to messages in a way that will undermine (or break) encryption, using untested (and as yet undefined) technology, and involve wide scale changes to existing practices. End-to-end encryption is the gold standard for data security and privacy across industry—it currently protects billions of people and businesses from a wide range of harms including data theft and cyber-attacks. Users access private messaging services in the knowledge and expectation that only they and the person or people that they are talking to can see the

[7] p46, 'Online Harms White Paper—Full Government Response', HM Government, 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944310/Online_Harms_White_Paper_Full_Government_Response_to_the_consultation_CP_354_CCS001_CCS1220695430-001__V2.pdf

content of their messages. We believe the Online Safety Framework should sit alongside the GDPR and the UK's Data Protection Act 2018 in supporting the protection of people's personal data.

Finally, the power itself is vague and unclear, leaving many important questions unanswered. For example, it is not clear what factors Ofcom will take into account when making "operational decisions" about whether a service should be required to use an accurate technology. There is no detail on how Ofcom will assess a tool as "sufficiently accurate" before mandating it for use. While there are some safeguards built into these provisions, such as requiring public notices about its use, these are very limited and not in any way commensurate with similar powers established in other regimes – such as judicial oversight, ability to appeal, and an explicit requirement to consider the privacy impacts of any 'use of technology' notice (including the public interest in the integrity and security of the underlying services).

Recommendation: The Government should remove the technology notice provisions entirely as they are contrary to the 'systems-focused' approach of the Bill and threaten user privacy.

Risks assessments

A central part of Ofcom's mechanism for ensuring that in-scope services are meeting their duties will be through the regular risk assessments provided for in the Bill. Facebook supports the use of risk assessments to enable the intended systems and risk-based approach. However, we are concerned that by placing too much detail about risk assessments on the face of the Bill, the Government is inadvertently reducing Ofcom's effectiveness

The relevant duties of care in the draft Bill require that a risk assessment be undertaken for each of the duties, and that such assessments form the basis of the measures that a service takes to address these risks. However, based on the current draft text, the number of risk assessments and the level of detail and implied frequency required may inadvertently undermine the Government's policy goals by making services' actions less effective and putting the framework at risk of quickly becoming outdated.

For example, the child-related duties and adult-related duties are separate sets of duties that apply to a service in addition to the illegal content duties. However, there are potential overlaps in respect of the content covered by these duties, and the steps taken to comply with them. It is possible that a particular type of content could be subject to both the child-related duties and the adult-related duties. Because it is unclear which should take precedence, this suggests that providers will need to bear in mind the possibility of overlapping duties when taking steps to comply with their duties, and will need to design systems and processes that can handle content subject to overlapping duties without imposing conflicting requirements.

Facebook supports the concept of requiring periodic risk assessments to deal with particular harms. But the final framework needs to balance the task of giving overall guidance on the structure and

content of a risk assessment, with ensuring sufficient room for different approaches to different services or parts of a service, and leaving flexibility for the future as technology evolves. Primary legislation is not a suitable vehicle for this.

The draft Bill also mandates that services conduct their first assessments within 3 months of the initial publication of Ofcom's guidance. The risk assessments must be kept up to date and repeated before any "*significant change*" to a relevant aspect of the service's design or operation. The Bill does not specify how to conduct a risk assessment, how to determine whether a change qualifies as "*significant*", or how often a risk assessment needs to be updated. We look forward to further clarity on these issues in the guidance, but caution that 3 months is a very tight timeframe within which to complete all relevant risk assessments in scope for the first time. Services can also contain multiple discrete elements or products and, as written, the draft Bill is unclear about how the boundaries between a service or its constituent elements are to be drawn. For this reason, such details should be left to Ofcom to determine.

In addition, if the definition of "significant change" is too broad and includes testing of new features, there is a risk that services may have to update and re-issue risk assessments in an almost continuous loop. This could have a detrimental effect on innovative proposals to test or launch new features, including features that have as their aim to *reduce* harms, which form a substantial proportion of the new features that we develop and release on our services. We hope that due regard will be given to striking the appropriate balance to promoting the development, testing, and launch of these integrity-focused products.

As with transparency requirements, new regulations in other jurisdictions also plan to take a risk assessment approach, for example, the proposed EU Digital Services Act. Through the Online Safety Bill, the UK has an opportunity to set standards and best practices globally by making use of Ofcom's existing connections with regulators and services internationally, as well as to align requirements in a way that reduces burdens on the tech sector as a whole.

Recommendation: The DCMS Committee should encourage Ofcom to seize the opportunity to set global standards and best practices for risk assessments using its existing connections with regulators. The final Bill's provisions should give overall guidance on the structure and content of the risk assessments, while ensuring sufficient flexibility for different services and adaptability for the future. This should include proportionate timetables, and protections for service changes that are themselves intended to reduce harm. Operational details should be set out by Ofcom in a format that can be updated, rather than written on the face of the Bill.

Checks and Balances

It is important that Parliament and the public have the opportunity to consider and discuss what is an appropriate balance of powers between Parliament, the Secretary of State, and the Regulator when it

comes to regulating online speech—especially when the intention is to regulate legal speech. The current draft framework places significant power into the hands of the Secretary of State to intervene in the regulatory framework after it has been passed by Parliament. These range from powers to give guidance to Ofcom,[8] through to the power to issue directions to Ofcom in some situations.[9]

Facebook's view is that this discussion needs to take place informed by further detail from the Government about how and where the different powers granted to the Secretary of State can be used. For example, what will be the role played by the Government's 'Objectives' for the Online Safety Codes of Practice set out in Clause 30? Under what circumstances might Ministers amend the Objectives as they are empowered to in Clause 30(5)? On what basis might ministers reject draft Codes of Practice, as provided in Clause 32? What is the definition of the term 'in line with Government Policy' as a reason for Ministers to give directions to Ofcom to modify a code, in Clause 33?

As well as providing answers to these questions, the Government should ensure that there is a formalised role for external input from experts in developing codes of practice, as has been the case with the New Zealand Netsafe code, and as is the case in Facebook's development of our own Community Standards.

Recommendation: The DCMS Committee should carefully scrutinise the Secretary of State's powers to overrule Ofcom, such as to direct changes to Codes of Practice, and ensure that these are sufficiently clearly described and, if appropriate, subject to Parliamentary oversight and input from civil society organisations. The UK should draw inspiration from domestic and international examples of how to balance powers when regulating speech, and make use of external and civil society expertise where appropriate.

4. International consistency

Q What are the lessons that the Government should learn when directly comparing the draft Bill to existing and proposed legislation around the world?

Alignment

Facebook wants to be a constructive partner to governments globally, as they weigh the most effective, democratic, and workable approaches to address online content.

[8] Clause 113(1)

[9] Clause 33(1), Clause 112(2) and (3)

Global content regulation varies widely, and we believe that certain regulations are more effective in addressing harms than others. For example, practically, we agree that there are lessons to be learnt from NetzDG. It is understandable that policymakers want to hold digital companies to be accountable, but an approach that is based on a single piece of content is difficult as it creates the thin line between freedom of speech and forbidden speech on the other side.

The primary weakness of the NetzDG system is that it is based on the obligation to assess all reported content under NetzDG within short timeframes: 24 hours for obviously illegal content and 1 week for less obviously illegal content. This system does not make a difference between more dangerous or less dangerous content, rather, all content has to be looked at within 24 hours. **As a result, a less brutal insult that has 4 views has to be treated equally to a live IS terror video that has 3 million views.** As a company, we could priortise certain more harmful and more dangerous content types first, e.g. terrorist material or child abuse before looking at more harmless content types. However, NetzDG does not allow for this flexibility.

Policymakers should take an active decision about what their main goal is: is it that platforms should effectively and quickly take action against harmful content? In that case our experience suggests that a system focused on the enforcement of our house rules will be most effective. Or is it more important that we assess reported content under national law? This takes much more time, and is thus much less effective. You cannot have both - effective and quick action and a correct legal assessment under national law done by an international company under short timeframes. Further, research on the NetzDG law has shown that the law has not been as effective as companies have been moving more towards proactive detection of harmful content, and it has created an incentive for overblocking. For this reason we welcome the Draft Bill's approach to consider systems over individual pieces of content as a better way to ensure an appropriate balancing of safety, freedom of expression.

As a comparison, we are working with Netsafe New Zealand, with regard to the development of a voluntary code of practice for online safety and harms. NetSafe describes the basis of the Code as 'trust through transparency'. The Code will include areas such online harms specific to children and youth, bullying and harassment, hate speech, incitement of violence, misinformation and disinformation, but specifically excludes areas that are already clearly already covered by legislation (for example, CVE or CEI). Similar to the UK, Netsafe New Zealand considers systems and processes, rather than incorporating simple notice and take down regimes for specific pieces of content.

We want a legal framework that allows us to tackle the problem of keeping the Internet safe and play our part in creating a healthier online environment, while safeguarding free expression and the benefits that the internet brings to people in the UK and around the world. The UK has an opportunity through the Online Safety Bill to take a truly international approach, building on the work underway with the EU's Digital Services Act to create efficient and effective regulation, which in turn will foster digital growth. There is no longer a separate digital economy and traditional economy in the UK, Europe or beyond – digital tools are at the heart of how every sector and every organisation operates. It is therefore worth the committee giving consideration to the ongoing development of the EU's Digital

Services Act and ensuring that the UK's Online Safety Bill doesn't inadvertently create opposed or conflicting obligations, particularly as both pieces of legislation share the same aims.

Recommendation: The Government should seek international alignment where it has been proven to be effective, rather than asking companies both small and large - who operate on a global scale - to undertake the significant task of adopting various regulatory approaches to addressing harmful content.

List of Recommendations

Overall Objectives
Systems-focused Approach: The DCMS Committee should encourage the Government to ensure that the final Bill aligns fully with the systems- and process-based framework which it has developed with stakeholders since 2017. The proposed definition of harm and certain provisions in the Draft Bill as written risk derailing the whole approach towards a focus on individual pieces of content and should be revised to ensure overall coherence and effectiveness. Further details on implementing the framework should be left to Ofcom.
Platform Design and User Agency: The DCMS Committee should continue to champion the importance of user agency and media literacy as integral parts of the UK's online safety landscape, and should encourage Ofcom and DCMS to move ahead with implementing the Online Media Literacy Strategy while the Bill is passing through Parliament.
Tensions, contradictions and workability
Tensions and Contradictions: The DCMS Committee should call on the Government to fully define the conflicting duties placed on services in the Bill, and provide further clarity on how services will be expected to balance them, either through draft Secondary Legislation or by setting out how Ofcom will evaluate services for striking these balances.
Democratic and Journalistic Content: The provisions relating to journalistic and democratic content should be removed entirely from the Bill, as they add additional risks to the framework rather than reducing them. If they are retained, the Government must set out clearly defined parameters around the definitions of all content types, and the connected duties, so that in-scope services have greater clarity on their responsibilities.
Fraud: Instead of tasking Ofcom to address online fraud through the Online Safety Bill, the

FACEBOOK

Government should involve Ofcom in the ongoing cross-Government work to tackle fraud, including the Fraud Action Plan and the Online Advertising Programme. Ofcom and the financial regulators should work together to ensure any proposed steps to tackle fraud through user-generated content are workable and align with Ofcom's OSB obligations.
Categorisation of Services: The Bill should set a more sophisticated method of defining 'Category 1' services, that looks not just at size and functionality, but at presence of harms and mitigating actions taken by the service. Ofcom should be given additional powers to carry out research into the prevalence of legal harms on small services, and the power to bring services into the scope of Category 1 for a period even if they do not meet the necessary thresholds. Details should be provided much earlier about any thresholds to give providers of different services clarity sooner.
Private Messaging: If no distinction is to be made between public and private online services, then private communications should be removed from the scope of the Bill entirely. Alternatively, the Government should revert to its original approach and set out differentiated obligations on services based on their privacy or accessibility, and taking account of relevant technical capabilities.
'Use of Technology' Notices: The Government should remove the 'use of technology' notice provisions entirely as they are contrary to the 'systems-focused' approach of the Bill and threaten user privacy.
Risk Assessments: The DCMS Committee should encourage Ofcom to seize the opportunity to set global standards and best practices for risk assessments using its existing connections with regulators. The final Bill's provisions should give overall guidance on the structure and content of the risk assessments, while ensuring sufficient flexibility for different services and adaptability for the future. This should include proportionate timetables, and protections for service changes that are themselves intended to reduce harm. Operational details should be set out by Ofcom in a format that can be updated, rather than written on the face of the Bill.
Checks and Balances: The DCMS Committee should carefully scrutinise the Secretary of State's powers to overrule Ofcom, such as to direct changes to Codes of Practice, and ensure that these are sufficiently clearly described and, if appropriate, subject to Parliamentary oversight and input from civil society organisations. The UK should draw inspiration from domestic and international examples of how to balance powers when regulating speech, and make use of external and civil society expertise where appropriate.

FACEBOOK

International Consistency
Alignment: The Government should seek international alignment where it has been proven to be effective, rather than asking companies both small and large - who operate on a global scale - to undertake the significant task of adopting various regulatory approaches to addressing harmful content.