Communications and Digital Committee: Freedom of Expression Online
Global Partners Digital Submission

GLOBAL PARTNERS DIGITAL
January 2021

About Global Partners Digital

Global Partners Digital (GPD) is a social purpose company working to enable a digital environment underpinned by human rights. We welcome the Committee's inquiry into freedom of expression online and were pleased to be able to provide oral evidence to the Committee.

In this submission, we focus on a number of the questions posed by the Committee relating to how the right to freedom of expression is impacted both by online platforms' content moderation policies and by government regulation. We have analysed and answered these questions on the basis of the right to freedom of expression as protected under the international and European human rights frameworks.[1] Where relevant, we also review the proposals in the response to the government's Online Harms White Paper, published in December 2020, given its relevance to the Committee's inquiry.

Question 1: Is freedom of expression under threat online? If so, how does this impact individuals differently, and why? Are there differences between exercising the freedom of expression online versus offline?

Addressing the third part of the question first, there are two distinct components of the right to freedom of expression requiring consideration: the scope of the expression which is protected, and the means by which that expression is manifested.

In relation to the scope of expression protected, there is no difference under either international or European human rights law between what expression is protected online as opposed to offline. Under the International Covenant on Civil and Political Rights (ICCPR), for example, the right to freedom of expression is guaranteed regardless of the "media" used by the person and makes no distinction between different media.[2] The UN Human Rights

[1] Primarily via Article 19 of the International Covenant on Civil and Political Rights (ICCPR) and Article 10 of the European Convention on Human Rights (incorporated into UK law via the Human Rights Act 1998).
[2] Article 19(2) of the ICCPR.

Committee has made clear that the right includes "electronic and internet-based modes of expression".[3] As such, that which we are free to seek, receive and impart in the offline world, we should be equally free to seek, receive and impart online, a position which has been confirmed by the Human Rights Council4 and the UN Special Rapporteur on freedom of expression.[5]

In relation to the *means* by which the right to freedom of expression is exercised online there are some differences as to how that right is exercised offline. For example, it is much easier to express oneself in a way that will reach larger numbers of people online than it is offline, it is easier to direct one's expression towards a particular individual, and it is easier to speak anonymously. For the most part, these are all positive and empowering aspects which have strengthened individuals' rights to freedom of expression, however they have also created new challenges which online platforms and governments are seeking to address.

The most significant difference, perhaps, is that the exercise of individuals' right to freedom of expression online largely relies upon spaces and channels provided by social media, search engines and messaging platforms, all of which are owned by private companies. As such, the private sector, and a small number of particularly powerful companies, develop and enforce many of the rules of what individuals can and cannot say or do online. In the offline environment, this role is predominantly (although not exclusively) undertaken by governments, where the rules are set out in legislation, and enforced through law enforcement agencies and courts.

Addressing the first part of the question, as the arbiters of what we can and cannot say and do online, the main risks to freedom of expression online stem from the actions of these online platforms (and in particular the content moderation policies developed and implemented by online platforms) and by regulation of online speech developed by governments. This latter category includes both restrictions which target individual speakers, and, increasingly, regulation of the online platforms themselves and their content moderation policies.

In relation to the actions of online platforms, threats to freedom of expression take many forms:

- Content moderation policies which are more restrictive than international human rights law and standards would permit;
- Vague or broadly worded content moderation policies that are unclear to users and leave significant discretion in their enforcement;
- The use of automated processes and AI to identify and moderate content, often despite high rates of inaccuracy and disproportionate impacts upon marginalised communities;
- Content moderators who have limited understanding of the language and context of the content that they are reviewing, and short time periods in which a decision is expected to be made;
- Inconsistent application of content moderation policies that leave some groups and individuals unprotected; and

[3] UN Human Rights Committee, General Comment No. 34: Article 19: Freedoms of opinion and expression, UN Doc. CCPR/C/GC/34, 12 September 2011, Para 12.
[4] See, for example, UN Human Rights Council Resolution 38/7, "The promotion, protection and enjoyment of human rights on the Internet", UN Doc. A/HRC/RES/38/7, 17 July 2018, which states that "the same rights that people have offline must also be protected online, in particular freedom of expression, which is applicable regardless of frontiers and through any media of one's choice".
[5] See, for example, UN Human Rights Council, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression", UN Doc. A/HRC/38/35, 6 April 2018, Para 1.

- An absence of effective appeal and remedial mechanisms which enable users to challenge decisions and obtain an effective remedy where a decision is upheld.

In relation to regulation by government, threats to freedom of expression online can take the form of overly restrictive criminal and civil laws prohibiting certain forms of speech, whether offline or online. Increasingly, however threats come from regulation of online platforms themselves, and the creation of regulatory frameworks which strongly incentivise the removal of content which may *potentially* be illegal, or even merely harmful but perfectly legal, with no countervailing incentives to ensure that the right to freedom of expression is not undermined, nor sufficient safeguards to mitigate this risk. Unfortunately, the UK government's proposals for a new Online Safety Bill would pose significant risks to the exercise of the right to freedom of expression in a number of ways.

The duty of care model

The proposed model is to impose upon online platforms a "duty of care" to take reasonable steps to protect users from harm. This model goes beyond simply ensuring that users are able to report content which is illegal or harmful so that it can be removed, but also taking steps to prevent users from coming across that content in the first place.6 As such, the model proposed implies a preventative approach rather than a reactive one.

There are two main ways that online platforms could prevent users from coming across harmful content which would be consistent with this approach. The first is to prevent it from every being made available on the platforms through checking all content beforehand; the second is to proactively and continuously monitor all content on the platform and remove harmful content as soon as it is identified with the hope that it will not have been seen.

Were the equivalent measures proposed in the offline world, they would be terrifying and unquestionably violations of the right to freedom of expression. The first is equivalent to requiring all individuals in the UK to have what they would like to say approved before they can say it, in case they wish to say something illegal or harmful. The second is equivalent to having everything anyone in the UK ever said monitored in case it is illegal or harmful. Such proposals should, without question, be considered disproportionate ways of addressing illegal and harmful expression in the offline environment, and this is no less true simply because they are being proposed in relation to what is said online. Indeed, the UN Special Rapporteur has stated that "states (...) should refrain from establishing laws or arrangements that would require the "proactive" monitoring or filtering of content, which is both inconsistent with the right to privacy and likely to amount to pre-publication censorship".[7]

Turning online platforms into arbiters of legality

The proposals would require online platforms to make determinations over whether content was legal or illegal. While there is some online content which a reasonable person could safely assess as likely to be illegal, making determinations over legality in the vast majority of cases requires time, investigation and legal expertise. High-profile cases relating to online content have shown that even the courts have found it difficult to draw the lines as to what is legal or illegal, with convictions by lower courts overturned.[8] There would be no equivalent to the

[6] See, for example, Para 35 of the government's response to the Online Harms White Paper which states that the duty of care will require companies to "prevent the proliferation of illegal content and activity online."

[7] See above, note 5, Para 67.

[8] See, for example, *Chambers v Director of Public Prosecutions* [2012] EWHC 2157 (Admin) (in relation to an allegedly "menacing" tweet, prosecuted under section 127(1)(a) of the Communications Act 2003) and *Scottow v Crown Prosecution Service* [2020] EWHC 3421 (Admin) (in relation to "causing

Human Rights Act 1998 which at least requires public authorities that enforce the law such as the police, the CPS and the courts, to exercise their duties in a way which complies with human rights, including the right to freedom of expression. Forcing companies to make decisions over legality without the necessary time, ability to investigate, and legal expertise risks removal of perfectly legal content.

Exacerbating this risk is the fact that there would only be sanctions where not enough content was removed, as opposed to too much, creating a strong incentive to remove "grey area" content rather than risk sanctions. Evidence from the implementation of the Network Enforcement Act in Germany in 2018 suggests that this would likely be the case: since the introduction of the law which requires companies to remove "manifestly unlawful" content in 48 hours or face heavy fines, there have been a number of instances of online platforms such as Twitter and Facebook removing pieces of content which were controversial, satirical and ironic, but not obviously illegal or even harmful.[9] It is for reasons such as these that the UN Special Rapporteur has said that governments "should avoid delegating responsibility to companies as adjudicators of content, which empowers corporate judgment over human rights values to the detriment of users".[10]

Automated processes

Given the scale of content which online platforms would need to review, the proposals would strongly incentive the use of automated processes to identify and remove illegal or harmful content, despite significant evidence that such processes are often inaccurate and unable to consider context, meaning large quantities of perfectly legal content would be swept up.[11] Indeed, the White Paper repeatedly refers to the use of AI to tackle certain forms of illegal and harmful content. AI is, however, at a very nascent stage when it comes to analysing speech and can only accurately identify a very small number of categories of speech which don't require an assessment of context or other nuances. As such, there are particular risks to freedom of expression which stem from the use of automated processes in order to determine whether content is illegal or harmful.

For one thing, it is simply not possible to develop accurate automated processing to identify particular forms of "illegal" or "harmful" content if the definitions of those forms of content are not clear. As such, automated processing will lead to inaccurate results and either the removal of legal and/or harmless content, or a failure to remove to illegal and harmful content.

For another, as noted above, making a decision about whether a particular piece of content is illegal or harmful requires an understanding of the context; however, automated processes are unable to determine context (or factors such as sarcasm, satire or irony). For example, it is impossible to know without context whether an online post which simply states "I'll see you in Shoreditch on Friday. Be ready!" is threatening violence, or simply a friend arranging to see another. An automated process could easily identify such a statement as a threat of violence and either remove it or prevent it from being uploaded at all. A video of violent and graphic war crimes could be terrorist propaganda or important evidence shared by human rights defenders. An automated process would not be able to tell the difference.

---

annoyance, inconvenience or needless anxiety to another [by] persistently mak[ing] use of a public electronic network" under section 127(2)(c) of the Communications Act 2003).

[9] See, for example, Scott, M. and Delcker, J., "Free speech vs. censorship in Germany", *Politico*, 14 January 2018, available at: https://www.politico.eu/article/germany-hate-speech-netzdg-facebook-youtube-google-twitter-free-speech, and Kinstler, L., "Germany's Attempt to Fix Facebook Is Backfiring", The Atlantic, 18 May 2018, available at: https://www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/.

[10] See above, note 5, Para 67.

[11] See, for example, Center for Democracy & Technology, "Mixed Messages? The Limits of Automated Social Media Content Analysis", 28 November 2017.

Legal but harmful content

The proposals would create strong pressure on companies to remove "legal but harmful" content. While the proposals focus on lawful content which risks causing "a significant adverse physical or psychological impact on individuals", there are significant concerns in principle with the use of regulatory pressure being used to restrict certain forms of expression online which would be lawful if expressed offline, to which we turn later in this submission.

Risks to encryption and privacy

The proposals would apply to private and encrypted channels, such as WhatsApp and Signal, which are relied upon by vulnerable groups, journalists, human rights defenders and others who would otherwise face persecution if they spoke openly about certain subjects, or if their communications could be accessed by third parties. As such, their very ability to exercise the right to freedom of expression in certain circumstances necessitates the privacy and security that encryption and other privacy-enhancing tools provide. If the obligation to monitor content on their platforms to identify illegal or harmful content extends to these private and encrypted channels, compliance would be impossible unless they were no longer private and encrypted, putting those users who use them at risk, or forcing them into silence.

Risks of copycat legislation being adopted elsewhere

Finally, while the focus of this inquiry is the UK, the internet is global in nature, and we wish to highlight the international dimension to this issue. There has been a recent trend of states passing copycat legislation relating to the internet, including that regulating online content. For example, shortly after the introduction of the Network Enforcement Act in Germany, a near-identical version was put forward in the Russian Duma.[12] However, while there are certainly concerns in relation to the German legislation, the adoption of the legislation in Russia would be even more problematic given the absence of any effective national human rights framework and the existence of criminal laws which prohibit expression in violation of international human rights standards.

As such, any proposals which are put forward in the UK, have the potential to be adopted in other states which could then point to the UK framework for justification. In states where speech which should be protected under international human rights law is criminalised or where there are no effective safeguards, such as an independent judiciary or a national human rights institution, for example, the effects could be far more restrictive than they would be in the UK. This would be hugely damaging for the UK's reputation as a strong proponent of freedom of expression and a free, open and secure internet.

[12] Reporters Without Borders, "Russian bill is copy-and-paste of Germany's hate speech law", *rsf.org*, 19 July 2017, available at: https://rsf.org/en/news/russian-bill-copy-and-paste-germanys-hate-speech-law.

Question 3: Is online user-generated content covered adequately by existing law and, if so, is the law adequately enforced? Should 'lawful but harmful' online content also be regulated?

Across the UK there is a significant amount of legislation which prohibits certain forms of expression or activity, and which apply online as much as offline. Many of these were listed in the government's Online Harms White Paper and include criminal offence relating to child sexual exploitation and abuse, terrorist content, extreme pornography, sharing "private sexual photographs and films", harassment, and the incitement of hatred and violence. Other non-criminal laws prohibiting certain forms of speech and behaviour, such as defamation and harassment also apply online as much as offline.

As well as these general speech-related offences, are also offences which apply specifically to the online environment in section 1 of the Malicious Communications Act 1988 (sending electronic communications with the intention of causing distress or anxiety) and section 127 of the Communications Act 2003 (sending by means of a public electronic communications network a message or other matter that is grossly offensive or of an indecent, obscene or menacing character).

We believe that these laws provide a comprehensive and robust framework setting out what is prohibited online, and addressing societal harms. Indeed, if anything, the existing law is already too restrictive from a freedom of expression perspective. In its recent consultation paper on "Harmful Online Communications", the Law Commission noted its concern relating that section 1 of the Malicious Communications Act 1988 and section 127 of the Communications Act 2003 "are sufficiently broad that they could, in certain circumstances, constitute a disproportionate interference in the right to freedom of expression".[13]

The challenge is not identifying which further forms of expression should be restricted, whether online or offline, but how to make sure that the existing law is properly enforced. This means, in particular, ensuring that the appropriate agencies, including the police, the CPS and the courts, have the resources needed to ensure that criminal activity online is investigated, prosecuted and that those found guilty are appropriately punished. While the nature of online communications may raise challenges for these agencies, this has always been the case where new technologies have developed, and the answer is to ensure sufficient resources and support are available, rather than to outsource law enforcement to online platforms. Unfortunately, this is what the government's proposals risk doing, by forcing online platforms to make decisions around the legality of online content. Doing so does a disservice to those who are the victims of criminal activity online: instead of the offender being brought to justice and appropriately sentenced by the courts, there will be no consequences for them save the deletion of the relevant content.

As such, we do not believe that additional forms of speech or behaviour which are currently lawful but considered "harmful" should be regulated when they occur online, but not offline. It would be wrong from a freedom of expression perspective to have identical forms of speech permitted offline but not online. If the government or Parliament considers that there is a particular form of speech or behaviour that should be permitted, the appropriate response is to prohibit it through clear, precise legislation which applies offline and online, and for that law to be enforced through appropriate public bodies. We note that the UK government's proposals are that where online content is not illegal, but there is nonetheless a risk that it

[13] Law Commission, Harmful Online Communications: The Criminal Offences: A Consultation paper, September 2020, Para 1.5.

would cause "a significant adverse physical or psychological impact on individuals", it would be considered as "harmful" and thus within the scope of the legislation. We would challenge the government to explain why, if a person does something which creates a risk of "a significant adverse physical or psychological impact on individuals", that behaviour is not illegal and why it would only be addressed through regulation when it occurs online but not offline.

Question 4: Should online platforms be under a legal duty to protect freedom of expression?

Yes. While international human rights law does not bind companies, the United Nations Guiding Principles on Business and Human Rights, the universally accepted business and human rights framework, make it clear that states have an obligation to ensure that human rights are respected by companies.[14] In the UK, for example, companies are bound by various legal obligations and duties which, in practice, ensure that they protect human rights. These include the Equality Act 2010, the Corporate Manslaughter and Corporate Homicide Act 2007, the Health and Safety at Work etc Act 1974, and the Modern Slavery Act 2015. Together, these ensure that companies have legal duties to protect, among other things, the right to life, the right to security, the prohibition on slavery, and the right to non-discrimination. However, there is currently no legislation which requires companies to protect freedom of expression, despite the fact that online platforms play a significant and ever-increasing role in determining how that right is exercised. Just as there are now legal duties upon companies in the UK which help ensure that they respect a wide range of human rights, we believe the time is ripe for further duties applying to online platforms to ensure that the right to freedom of expression is similarly respected.

A legal duty would help ensure that online platforms take a human rights-based approach to the issue of content moderation, ensuring that their policies are not overly restrictive but in line with international human rights law and standards, that they are implemented in a way which mitigates risks to freedom of expression, are applied in a non-discriminatory manner, and that there is greater transparency over decisionmaking.

Importantly, a legal duty on online platforms to respect freedom of expression would not stop them from taking appropriate decisions to restrict content where it served a legitimate purpose, just as the legal obligation on governments under international human rights law does not prevent governments from prohibiting certain forms of speech where it is a proportionate means of achieving a legitimate objective.

Question 5: What model of legal liability for content is most appropriate for online platforms?

We believe that the current model of legal liability for content is the most appropriate one, namely that which derives from Article 14 EU's E-Commerce Directive. Under this model, an online platform is not generally liable for the content that it hosts. However, it can become liable once it becomes aware of illegal activity or information and does not act "expeditiously" to remove that illegal information. Article 14 is also clear that courts and other administrative bodies can require online platforms to remove content which it has determined is illegal.

[14] UN Guiding Principles on Business and Human Rights, Principle 3.

When the EU Commission published its proposals for a new Digital Services Act in December 2020, it explicitly decided to retain its existing intermediary liability regime. The UK government also announced in its Online Harms White Paper that the proposals would be "compatible with the EU's e-Commerce Directive, which limits their liability for illegal content until they have knowledge of its existence and have failed to remove it from their services in good time".[15] We strongly support this decision. To hold online platforms liable for the content hosted - as is the case in some countries, such as China and Thailand - would likely lead to online platforms either withdrawing their services from the UK so as to avoid legal liability for the activities of their users, or introducing highly risk averse and censorious content moderation policies and processes. Both of these options would have grave impacts upon the right to freedom of expression.

Question 6: To what extent should users be allowed anonymity online?

The ability to exercise anonymity online, including when searching for information and expressing oneself, can be critical for the enjoyment of human rights, particularly the right to respect for private life (which includes the ability to correspond with others privately) and to freedom of expression; indeed, it is only through anonymity that many in society - such as journalists, whistleblowers, human rights defenders and persecuted minority groups - are able to express themselves safely online. In his 2015 report to the UN Human Rights Council, the then UN Special Rapporteur on freedom of expression confirmed that the ability to remain anonymous online was protected by both of these human rights.[16]

Any restrictions on the ability of an individual to be anonymous online must therefore comply with the requirements of international human rights law, i.e. they must be provided for by law; may only be imposed for legitimate grounds; and must conform to the strict tests of necessity and proportionality. Further details on what this requires in practice were also set out in the UN Special Rapporteur's report.

To comply with the first requirement, any restriction on anonymity must be "precise, public and transparent" and must "avoid providing State authorities with unbounded discretion to apply the limitation".[17] Furthermore, "strong procedural and judicial safeguards should also be applied to guarantee the due process rights of any individual whose use of encryption or anonymity is subject to restriction" and, in particular "a court, tribunal or other independent adjudicatory body must supervise the application of the restriction".[18]

To comply with the second requirement, anonymity should only be restricted in order to protect specified legitimate interests, namely the rights or reputations of others; national security; public order; public health or morals.[19] And to comply with the third requirement, the restriction must be "necessary" to achieve the legitimate objective. The European Court of Human Rights has made clear that this means that it must be something more than "useful," "reasonable" or "desirable".[20] Any restrictions should be "subject to independent and impartial judicial authority, in particular to preserve the due process rights of individuals".[21] Furthermore, "necessity also implies an assessment of the proportionality of the measures

[15] HM Government, Online Harms White Paper, April 2019, Para 41.
[16] UN Human Rights Council, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc. A/HRC/29/32, 22 May 2015.
[17] Ibid., Para 32.
[18] Ibid.
[19] Ibid., Para 33.
[20] See, for example, The Sunday Times v. United Kingdom, judgement of 26 April 1979, Para 59
[21] See above, note 16, Para 34.

limiting the use of and access to security online" and so the restriction must be "the least intrusive instrument amongst those which might achieve the desired result".[22]

In practice, this means that the default position should be that individuals remain free to be anonymous online. If an individual is to be restricted from doing so on a particular platform, the situations in which that restriction is to be applied should be set out clearly and precisely in legislation, only permitted on a case-specific basis where it is a necessary and proportionate means of achieving a legitimate aim, and the restriction should be subject to judicial oversight and other safeguards.

Question 9: How could the transparency of algorithms used to censor or promote content, and the training and accountability of their creators, be improved? Should regulators play a role?

Greater transparency of algorithms which are used to moderate content (whether through removal, prioritisation or deprioritisation, or otherwise) is critically important to understanding how decisions are made which impact upon people's right to freedom of expression. At the very minimum, we believe that online platforms should be legally required to inform users when algorithms are used which impact upon the online content that they share or access. Online platforms should also be required to inform users what personal data and data points are used by the algorithms, and give users greater choice to choose between algorithms and to turn off the use of algorithms which affect content where feasible (for example, by viewing content simply chronologically on a social media platform's feed).

Regulators should play a role, namely to audit whether online platforms are complying with those legal duties. The regulator should have appropriate investigatory powers to inspect the algorithms used by online platforms or be provided with sufficient information. The regulator should also be empowered to publish information on its investigations and, where necessary and appropriate, make requirements of online platforms to ensure compliance.

Question 10: How can content moderation systems be improved? Are users of online platforms sufficiently able to appeal moderation decisions with which they disagree? What role should regulators play?

There are a number of ways that content moderation systems can be improved to ensure that freedom of expression is better protected. First, online platforms should develop terms of service and content moderation rules that are in line with international human rights standards, as recommended by the UN Special Rapporteur on freedom of expression. The terms of service and rules should be clear and precise so that users know what is and is not permitted on platforms, should only lead to restrictions on content where a legitimate aim is pursued (such as to prevent crime, the incitement of violence, or to protect other users' human rights), and should be implemented in a proportionate manner.

[22] Ibid.

Second, a human should always be involved in content moderation decisions. While automation may play a role in flagging content, a human should always be involved in the actual decision to remove or otherwise moderate content. That person should have sufficient knowledge of the relevant language and context in order to be able to make an informed decision.

Third, whenever content is removed or a user's account suspended, they should be provided with a clear and detailed explanation setting out the reasons for the decision. Users should always be able to challenge these decisions. Appeals should enable them to provide relevant context and information, and to be considered in a timely and transparent manner.

As with algorithmic transparency, regulators should play a role in relation to the development and enforcement of content moderation policies. The regulator should have appropriate powers to publish guidance on the development and enforcement of content moderation policies (including the ability to appeal decisions). The regulator should also have powers to inspect and audit the development and enforcement of content moderation policies or be provided with sufficient information. As with algorithmic transparency, the regulator should also be empowered to publish information on its investigations and, where necessary and appropriate, make requirements of online platforms to modify policies and practices to ensure compliance.

Question 12: Are there examples of successful public policy on freedom of expression online in other countries from which the UK could learn? What scope is there for further international collaboration?

There are some examples of strong public policies which support freedom of expression online. In Germany, the State Treaty on the modernisation of media legislation in Germany (Medienstaatsvertrag) was adopted in 2020, the first legislation in the world requiring algorithmic transparency from online platforms. Under the legislation's provisions on transparency, online intermediaries will be required to provide information about how their algorithms operate, including the criteria that determine how content is accessed and found and the key criteria that determine how content is aggregated, selected, presented and weighed.[23] In Brazil, legislation regulating intermediary liability explicitly requires a court order before an online platform must restrict particular content (outside of the enforcement of its own terms of service).[24]

[23] Algorithm Watch, "Germany's new media treaty demands that platforms explain algorithms and stop discriminating. Can it deliver?", 9 March 2020, available at: https://algorithmwatch.org/en/new-media-treaty-germany/.
[24] Marco Civil da Internet, Law No. 12,965, Articles 18–19.