# Demos—written evidence (FEO0092)

## House of Lords Communications and Digital Committee inquiry into Freedom of Expression Online

1. Demos is Britain's leading cross-party think tank, with a 25-year history of high quality research, policy innovation and thought leadership. Our priority is to bring ordinary citizens' voices into policy making. The Centre for the Analysis of Social Media (CASM) is Demos' dedicated digital research hub, where our work focuses on tech and society. Our 'Good Web Project' is currently examining how we can create a future for the internet that is compatible with liberal democracy and principles such as freedom of expression.[1] We are responding to this consultation as questions around how best to protect and promote freedom of expression are crucial to our work, which looks at: how citizens can be brought fully into democratic processes; how regulation of online spaces should balance freedoms with protection from harm; and how platform design can encourage free and positive expression. Our responses to this consultation draw on our existing body of research, policy and advocacy work and our wider expertise.

### *Q1. 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?*

2. Exercising freedom of expression online should have the same level of protection, and the same constraints, as doing so offline. However, the conditions and context in which online speech occurs can be very different to offline speech. The same expression made in an offline or online context could be subject to different levels of protection or constraint, as the difference in context, reach and potential audience may warrant different levels of action.

3. Far from being a neutral entity that simply facilitates people's expressions, the way platforms are designed and run affects the kinds of expressions which are made and those which are permitted. From moderation rules to character limits to content prioritisation systems, these choices not only affect what content remains on a site, but influence what expressions are made in the first place by users. These constraints are not inherently infringements on freedom of expression. Indeed, we submit that the existence of these constraints (which are in many cases necessary to the functioning of the platform and the prevention of serious harm) is too often weaponised under the guise of concerns about freedom of expression, such as Republicans in the US attacking social media platforms for 'censorship' after they took action against disinformation.[2]

4. However, the fact of platform mediation does present possible ways in which freedom of expression online *can* be under threat. The realisation of these threats occur to different degrees in different online spaces and political contexts around the world.

---

1 https://demos.co.uk/project/the-good-web-project/

2 https://www.politico.com/news/2020/10/26/censorship-conservatives-social-media-432643

5. **People's online expressions being prevented:** People may be prevented from expressing themselves altogether on a particular platform e.g. if they have their account suspended or blocked. They may also have particular or multiple instances of their expressions being taken down. That platforms have and exercise these powers is crucial to enable accounts to be removed which are inauthentic, fraudulent or dangerous. It is consistent with the protection of freedom of expression to have lower barriers for speech being restricted in these ways, which are minimally invasive and keep other avenues for expression available, than exist for criminal prosecution.

6. However, there will be cases where people or instances of their speech will be removed wrongly: e.g. where it is judged to meet a threshold of harm or to violate terms of service when it does not. There should be, for such instances, clear avenues of appeal and redress for users, who should be provided with the information as to why a decision was made, and be able to appeal and have their content reinstated if it was in fact an incorrect application of moderation.[3] This is particularly the case where automated systems are used: whenever algorithmic systems are deployed in content moderation, a decision must be made on how to set the balance between correctly moderating relevant content and preventing the wrongful moderation of irrelevant content (as prioritising one tends to reduce the success of the other). This does pose potential problems for freedom of expression: if people's speech is being removed systematically without cause. This is especially worrisome as minority groups are at greater risk of having their speech wrongfully moderated.[4][5]

7. This is not to say that automated systems should aim for zero errors in moderation, even at the cost of missing seriously harmful content. It does mean that the automated systems being used, the basis on which those systems are making decisions, the reasons and effects of those decisions, should be transparent and available for scrutiny by a regulator, and ideally by civil society at large.

8. **Information being deprioritised:** A crucial tool which platforms should be utilising to reduce the harms that can occur from content online without engaging in sweeping takedown regimes is the deprioritisation of certain kinds of content which are harmful. This is not without ramifications for freedom of expression: being able to access information is a crucial part of exercising the right, and 'shadowbanning', where an individual's content can be demoted so that it does not reach its audience, can also be a way of restricting that individual's speech. However, all content curation prioritises some things and deprioritises others: there is no 'neutral' way to curate content online. Hence we support transparent systems of content curation, based on promoting goods such as public health online, rather than those currently employed, which promote content for engagement rather than public good.[6]

---

3 https://demos.co.uk/project/everything-in-moderation-platforms-communities-and-users-in-a-healthy-online-environment/

4 https://www.internetlab.org.br/en/freedom-of-expression/drag-queens-and-artificial-intelligence-should-computers-decide-what-is-toxic-on-the-internet/

5 https://www.aclweb.org/anthology/P19-1163.pdf

6 https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

9. Legislation to prevent harmful content online should focus on these system-level changes, rather than individual pieces of content. The Government's proposed Online Safety bill focuses on platform responsibilities to implement systemic changes rather than responsibilities to remove or monitor every individual piece of content which falls under a certain category. We submit that this approach greatly reduces this risk of platforms interfering with freedom of expression online. The focus on systemic changes is likely to be more effective, while also not incentivising platforms to take down sweeping categories of content online.

10. **Access to information being blocked**: Targeted blocks, where particular sites or sources of information known to promote illegal activity are an important part of preventing online harms - where there is appropriate oversight of how these targeted blocks are employed, and what grounds are required to demand one. However, widespread blocks are a very significant concern for freedom of expression and freedom of opinion, particularly in countries where internet shutdowns are used regularly, allegedly to reduce the risk of violence. Recent examples include the shutdown during the Ugandan elections, and shutdowns in India under the guise of tackling disinformation and promoting order.[7] It is incumbent that, as the UK moves to regulate the online world, we emphasise the importance of not abusing blocking powers and include clear and adequate safeguards for any targeted blocking.

11. **Hostile conditions online making free expression unsafe for particular groups:** We understand freedom of expression as a positive, as well as a negative freedom: that is, if someone does not have the conditions which would enable them to meaningfully exercise their right, that is a constraint on that right. This conception should be applied with caution, as it can and has been deployed to claim that views which receive criticism online, or for which users face platform penalties, are being 'censored', even in cases where those actions or critiques are based on the harm caused by the speech.

12. However, there is extensive evidence of the 'silencing' effect of harassment and abuse on social media, particularly against women, and especially Black and minoritised women.[8] This is a concern for these online spaces which increasingly function (despite being privately controlled) as extensions of the public sphere: as spaces where individuals communicate with each other, build networks of solidarity, organise and communicate with their representatives, it is important for freedom of expression and democratic values that this be addressed. Those who are digitally excluded, either through financial barriers to accessing devices or data, or skills barriers to using digital platforms confidently and safely, also risk being deprived of the ability to engage in these online public spaces and access necessary information.[9]

13. Action should thus be taken by platforms to ensure these spaces are not silencing of marginalised groups. This should include necessary post-hoc

---

7 https://edition.cnn.com/2021/01/14/africa/uganda-vote-internet-shutdown-intl/index.html; https://www.dw.com/en/indias-internet-shutdowns-function-like-invisibility-cloaks/a-55572554

8 https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/; https://fixtheglitch.org/covid19/; https://ogbv.pollicy.org/report.pdf

9 https://www.goodthingsfoundation.org/research-publications/digital-nation-2020

measures, providing users with the ability to report, block, or restrict access to their content or what content they see. However, this should also involve changes to the design of platforms and how the systems are governed and employed to reduce the chances of this type of harassment continuing.

### *Q3. 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?*

14. The current law is not adequate. Individual instances of lawful but harmful content are not appropriate for government regulation (though we are not here counting 'regulating' currently lawful but harmful content as including making a subset of that content illegal if appropriate). However, we do consider it appropriate that a regulator should have the power to demand that platforms which provide online services take steps to reduce the general risk of legal but harmful content, including enforcing their own terms of service consistently and transparently.

### *Q4. Should online platforms be under a legal duty to protect freedom of expression?*

15. Online platforms should be under a legal duty to protect the rights of their users. However, a duty to protect freedom of expression by itself risks incentivising companies to be excessively lax in moderating and investing in redesigns to make certain kinds of harms less likely, for fear of seeming to violate that duty. Any such duty should also have accompanying duties to promote other rights, along with acknowledgements of what the appropriate limits on freedom of expression are.

### *Q5. What model of legal liability for content is most appropriate for online platforms?*

16. We support the duty of care model under which platforms have a legal duty to have systems in place to act on illegal and harmful content. Models which make online platforms legally liable for individual pieces of illegal content, such as could potentially happen in the US under a Section 230 revocation, or conversely, make it illegal to remove any legal content from a site as is being proposed in Poland,[10] have a higher risk of incentivising over-or under-removal of harmful content.

### *Q6. To what extent should users be allowed anonymity online?*

17. As we outline in our report *What's in a Name?,* anonymity should be understood as a relational concept: an individual cannot be 'anonymous' in the abstract, but anonymous *from* a particular individual or organisation. A user can be anonymous from other users of an online service, but have to provide identifiable information to the platform itself: or a user could be anonymous from a platform but able to be identified by law enforcement. When discussing

---

10 https://www.theguardian.com/world/2021/jan/14/poland-plans-to-make-censoring-of-social-media-accounts

how far users should be anonymous, it is important to distinguish who that anonymity is from.[11]

18. Being able to control one's identity online is a crucial way for people to exercise their freedoms and protect themselves where identification could lead to harm against them. However, anonymity can also mean it is harder to hold people online accountable where they have committed harm against others. A significant threat to people being anonymous online is the commercial aggregation of personal data - not only that which is provided by the individual, but data about an individual's online activity, which can be aggregated or brokered to enable personalisation of content and targeted advertising.

19. We have proposed a three-fold test for how anonymity should function online in liberal democracies, when solutions attempt to balance these aspects of anonymity:[12] 'Future solutions must: 1. Protect internet users' ability to choose anonymity online, and emphasise its importance in preserving freedom of expression. 2. Allow accountable institutions tasked with preserving security under a democratic mandate to exercise their powers effectively. 3. Ensure users are able to provide meaningful consent to any deanonymisation by third-parties.'[13]

***Q7. How can technology be used to help protect the freedom of expression?***

20. Access to technology is a prerequisite for people to be able to enjoy the freedoms of the internet, particularly freedom of expression and access to information. Technologies will enable people to communicate and access information even when they are blocked or at risk from repressive states are also crucial to protecting freedom of expression globally: tools such as end-to-encryption and VPNs which allow safe communication are vital for people to be able to exercise their rights even when they are under threat.

***Q8. How do the design and norms of platforms influence the freedom of expression? How can platforms create environments that reduce the propensity for online harms?***

21. The way that platforms are designed significantly influences the kind of behaviour that is encouraged, the ways that users interact with each other, and the kind of content that is likely to proliferate. Platform design can include behavioural 'nudges', which prompt a user to act or refrain from acting a certain way; increasing friction to reduce the likelihood of certain kinds of posts, using moderation systems to deprioritise or block users or content, and involving the online community in the policing of the space through user feedback and moderation. However, platform design should also focus on encouraging pro-social behaviour, not merely reducing anti-social behaviour: rewarding users who engage with a platform over a longer period of time, or improving a user's ability to impact the online spaces they exist in, can encourage greater online community-building.[14]

---

11 https://demos.co.uk/project/whats-in-a-name/
12 https://demos.co.uk/project/whats-in-a-name/
13 https://demos.co.uk/project/whats-in-a-name/
14 https://demos.co.uk/project/everything-in-moderation-platforms-communities-and-users-in-a-healthy-online-environment/

22. Currently platforms are in control of which systems they deploy, and choose those based on which will maximise revenue. Where harms are addressed, it is reactive and inconsistent. To better address online harms, platforms need to fundamentally adjust their systems of governance and user empowerment. Without the ability to fundamentally change the commercial incentives, regulation is needed to shift the metrics by which platforms measure their success or failure.

***Q9. 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?***

23. Currently the algorithms used and who has created them and how is not transparent. The effects of different algorithms being used; the rationale behind that decision; and the ways in which they have been designed and trained are regularly opaque. Some companies have chosen to make information public, either about the results of applying different moderation models, or in some cases making the code itself available to (limited) external scrutiny.[15] However, these steps are at the companies' discretion, and companies may take some useful steps while neglecting the others which are crucial to full accountability for how algorithms are affecting what content is visible or permitted. The regulator can play a crucial role in requiring: information from companies on the performance of their algorithms and how they are operating; as well as inspections and compulsory tests or audits of algorithmic systems by a regulator or independent third party.[16][17]

***Q10. 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?***

24. Content moderation systems could be improved through: **Transparency:** The terms of service, principles and processes of moderation, as outlined in our report *Everything in Moderation*[18]*,* should be clearly communicated to the average user: including clarity over permissible content and behaviour, over moderation practices; and over a user's individual experience of a moderation decision. Terms should be communicated in accessible language, and be clearly located on the platform itself, with users directed to it periodically to ensure ongoing informed consent. These terms should be consistently applied.

25. **Redress:** If action has been taken against a user or their content, there should be clear systems of redress in place to which a user can appeal, and have their appeal considered and responded to transparently.

26. **Empowerment:** Users play a crucial role in effective content moderation systems by acting as moderators or admins of online spaces. The civic labour that they perform should be recognised, incentivised, and rewarded, and

---

15 https://www.wired.com/story/tiktok-finally-explains-for-you-algorithm-works/
16 https://demos.co.uk/wp-content/uploads/2020/04/Algo-inspection-briefing.pdf
17 https://www.adalovelaceinstitute.org/our-work/themes/algorithm-accountability/
18 https://demos.co.uk/project/everything-in-moderation-platforms-communities-and-users-in-a-healthy-online-environment/

moderators should receive support for dealing with particular crises which significantly affect moderation (such as Covid-19).[19]

27. The increased errors from reliance on automated systems during the pandemic as moderators were furloughed demonstrated the necessity of human moderation systems.[20] The experience of human moderators employed by platforms: from having to work in Covid-unsecure conditions, to the extreme psychological stress of dealing with horrific content, shows that there is a need for tech platforms to be investing more in their human moderators, through improved psychological care and supporting home working.[21][22]

28. A regulator can play a role in demanding that companies provide evidence of the processes they have implemented to ensure the above, and provide evidence of the efficacy of different moderation systems or choices which a platform uses.

***Q.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?***

29. The EU Digital Services Act looks to enshrine freedom of expression online, to ensure against general monitoring obligations; that users whose content is removed should be informed about the removal, the reason for it and have avenues for redress, and that risks to freedom of expression should be included in risk assessments carried out by platforms. Although the DSA is still at draft stage, and its efficacy has not been tested, given its significance for internet regulation, as well as its promotion of rights, we would hope to see the UK collaborating closely with the EU on aligning regulatory regimes. The closer aligned regimes are internationally, the more credible the regime, the clearer what protections are in place is for users, and the more likely tech companies are to comply.

*15 January 2021*

19 https://demos.co.uk/project/everything-in-moderation-platforms-communities-and-users-in-a-healthy-online-environment/

20 https://www.reuters.com/article/us-health-coronavirus-google/social-media-giants-warn-of-ai-moderation-errors-as-coronavirus-empties-offices-idUSKBN2133BM

21 https://www.foxglove.org.uk/news/open-letter-from-content-moderators-re-pandemic

22 https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona