The Electronic Frontier Foundation and New America's Open Technology Institute are submitting this joint set of comments as part of the government of the United Kingdom's consultation on its Online Harms White Paper.

Both of our organizations understand and share the government's concerns about wanting to increase the safety of users on the internet. However, the proposal outlined in the White Paper is concerning as it is broad and lacks clarity and detail around how this framework would be operationalized in practice. In addition to our comments below, we encourage the government to further refine and clarify its approach, and to ensure that in doing so, the right to free expression is adequately safeguarded and supported.

We outline below our responses to Questions 1 - 3, 6 - 7, 9, 11, 13 - 14, and 18 of the Consultation Questions in the White Paper.

1. This government has committed to annual transparency reporting. Beyond the measures set out in this White Paper, should the government do more to build a culture of transparency, trust and accountability across industry and, if so, what?

Transparency reporting is a valuable mechanism for promoting accountability for internet platforms. Robust transparency reports by companies can help users understand how they can safely engage online and can help ensure that companies are respecting the fundamental human rights, including privacy and free expression, of all users.

The White Paper notes that the regulator will have the power to require that reporting addresses, among other things, "Measures and safeguards in place to uphold and protect fundamental rights, ensuring decisions to remove content, block and/or delete accounts are well- founded, especially when automated tools are used and that users have an effective route of appeal." The government should particularly seek to use transparency reports as a mechanism for ensuring that free expression is not being stifled, particularly in the case of marginalized communities and individuals who may be subjected to hateful attacks. In addition, given that internet platforms are working to ramp up their content moderation efforts, by hiring new human moderators as well as by increasingly relying on automated tools to flag and moderate content at scale, transparency around these human and automated-decision making practices are extremely important. It is also critical to ensure that there are robust appeals processes that enable users to meaningfully challenge decisions to remove or otherwise action their content and accounts. Increased transparency around platforms' appeals processes are also integral, as these data points can highlight how often companies erroneously removed user content, and whether these users were afforded effective remedies.

In order to maximize the effectiveness of transparency measures as a mechanism to promote accountability, companies should standardize their reporting, so data points can be effectively compared and contrasted. In addition, data points reported on in transparency reports should be granular enough that they provide insights into how much user speech is being removed or

otherwise shaped and impacted, and into what content moderation tools are being used. However, it should be noted that in some cases the granularity of reporting must be limited to avoid intrusions on users' privacy rights. In addition, requirements related to transparency reporting from multiple jurisdictions can create complications for platforms, especially if there is a low common denominator in terms of the structure of the required reporting in different jurisdictions. As a result, governments and companies should identify and strive to provide a level of transparency that is meaningful but respectful of user rights, and preferably a standardized and useful format that works across multiple jurisdictions.

There are currently a number of existing resources which can guide companies on how to strike this balance and demonstrate transparency and accountability around their content moderation practices. These include:

- The Santa Clara Principles on Transparency and Accountability in Content Moderation: The Santa Clara Principles outline minimum standards technology platforms must meet in order to provide adequate transparency and accountability around their efforts to take down user-generated content or suspend accounts that violate their rules. The principles advocate for greater transparency and accountability by focusing on three key demands -- comprehensive numbers detailing their content moderation activities, clear notice to affected users, and a robust appeals process.[1]
- New America's Open Technology Institute's Transparency Reporting Toolkit on Content Takedown Reporting: This edition of the Transparency Reporting Toolkit surveys 35 global internet and telecommunications companies on how they are reporting on different types of content takedowns (e.g. Terms of Service based content takedowns, copyright-related takedowns). The Toolkit also offers a set of best practices for transparency reporting on these various forms of content takedowns which aims to help companies make their transparency reports more standardized and meaningful.[2]
- The Electronic Frontier Foundation's Who Has Your Back report: An annual report that for the last two years has focused on censorship. The report assesses major tech companies' content moderation policies across six categories related to transparency around content takedowns and notice and appeals mechanisms.[3]
- The Ranking Digital Rights Corporate Accountability Index: The Corporate Accountability Index evaluates 24 of the world's most powerful internet, mobile, and telecommunications companies on their disclosed commitments and policies related to privacy, security, freedom of expression and governance. It includes specific

[1] "The Santa Clara Principles On Transparency and Accountability in Content Moderation," https://santaclaraprinciples.org/.
[2] Spandana Singh and Kevin Bankston, *The Transparency Reporting Toolkit: Content Takedown Reporting*, October 25, 2018, https://www.newamerica.org/oti/reports/transparency-reporting-toolkit-content-takedown-reporting/.
[3] Gennie Gebhart, *Who Has Your Back? Censorship Edition 2019*, June 12, 2019, https://www.eff.org/wp/who-has-your-back-2019#executive-summary.

benchmarks for transparency around content moderation practices and requests received to remove or restrict content or accounts.[4]

In addition, in order to foster a culture of transparency, the government should also provide transparency around how it engages with companies around content removal, including disclosing how many requests the government has submitted to various platforms for content removal through its agencies as well as through entities such as Internet Referral Units.

2. Should designated bodies be able to bring 'super complaints' to the regulator in specific and clearly evidenced circumstances? If your answer to question 2 is 'yes', in what circumstances should this happen?

Greater clarity is needed on what the responsibilities of super complainants would be. The current language in the White Paper does not clearly specify whether these responsibilities would include reporting illegal content, a failure of a company to meet duty of care, or other responsibilities.

Should super complainants be responsible for flagging illegal content, we suggest that the government observe the "Trusted Flagger" model that many major internet platforms, such as YouTube operate. Trusted Flaggers are typically individuals, government agencies, or non-profit or civil society organizations who have an expertise with a particular content policy area, or a particular geographic region, and who have demonstrated high accuracy rates when it comes to flagging content to a platform that violates the platform's content guidelines. These Trusted Flaggers can utilize a range of additional flagging benefits that often include a bulk-flagging tool and prioritized flag reviews.

Any proposed designated body should demonstrate a strong track record of accuracy, reliability and legitimacy when it comes to flagging content that violates a platform's content guidelines. In addition, designated bodies should also demonstrate expertise in a given policy area or region, and there should be multiple bodies that cover various types of online harms. The government must be mindful of the fact that there may be entities with potential conflicts of interest that may have strong incentives to be designated as super complainants, whereas some organizations working on important online harms work may not receive such designations. As a result, resources must be dedicated towards ensuring a diverse group of perspectives are covered by these designated bodies.

In addition, in the United States, the First Amendment strictly limits the extent to which the government can regulate the content of speech, including online speech, and the government must be careful not to interfere with the right of free expression. Although in the United Kingdom, the government has more ability to control online speech, it is nonetheless important

[4] Ranking Digital Rights, *2019 Ranking Digital Rights Corporate Accountability Index*, May 2019, https://rankingdigitalrights.org/index2019/

that any process involving the creation of a designated body that can bring special cases to a government regulator, should include robust safeguards for free expression and user privacy.

3. What, if any, other measures should the government consider for users who wish to raise concerns about specific pieces of harmful content or activity, and/or breaches of the duty of care?

From the point of view of the complainant, we believe that the mechanism laid out in the regulatory framework is sufficient.

One element that the paper profoundly underestimates, however, is how online complaint mechanisms themselves can be used for abuse. Abusers can organize to present a complaint in the worst light possible, or complain in such numbers, or so frequently, that an unjust takedown of a particular victim's account or their content may inevitably occur. The government needs to be mindful that their — or companies' — complaint system will be mis-used in this fashion. As a result, the government should include a safeguard against this in the regulatory framework, and build into its operation mitigations against such attacks.

We believe that such mitigations can only be maintained if the incentives are correctly aligned so that repeated abuse can result in consequences for the regulatory authority, and/or the organizations providing complaints or enforcing the resolution of such complaints. Part of that incentive system can be through transparency and an effective and open appeals process — see, for instance, the provisions of the *Santa Clara Principles.*[5] Examples of existing, comparable safeguards can be found in the U.S. Digital Millennium Copyright Act (DMCA). Although the DMCA is problematic in many respects, it does have some protections against abuse of the complaint process, including the requirement that complainants must affirm under penalty of perjury that they are authorized by the rightsholder to make such a complaint.

6. In developing a definition for private communications, what criteria should be considered?

When developing a definition for private communications, the UK government should be particularly mindful of platforms that offer encrypted communications and messaging services. Access to encryption is fundamental for safeguarding human rights in the digital age, including privacy and freedom of expression.

The UK government should not create any rules or regulations that require internet platforms and technology companies to pre-filter or otherwise scan or monitor private communications. In addition, no rules or regulations should mandate interference with technology companies' ability to offer messaging services that are end-to-end encrypted.

[5] See *Santa Clara Principles*, https://santaclaraprinciples.org/

Encryption fosters innovation, generates trust and confidence with consumers, and is a key component of data protection and cybersecurity today. Attempts to undermine encryption will result in rights-threatening harms for all users, especially for those who are part of vulnerable communities.

7. Which channels or forums that can be considered private should be in scope of the regulatory framework? What specific requirements might be appropriate to apply to private channels and forums in order to tackle online harms?

As previously discussed, access to encryption including for end-to-end encrypted messaging services is vital for safeguarding the rights to privacy and freedom of expression. As a result, no channels or forums that offer private encrypted communications or messaging services should be in scope of the regulatory framework.

9. What, if any, advice or support could the regulator provide to help businesses, particularly start-ups and SMEs, comply with the regulatory framework?

Vast amounts of research have indicated that new regulations regarding online content regulation and filtering often negatively impact startups and SMEs that are unable to grapple with this growing range of requirements both financially and in terms of human capital. In order to help startups and SMEs better comply with the proposed regulations, the regulator should facilitate workshops for these organizations on what is necessary for them to comply with the framework. These workshops should be held regularly, so new startups can also receive this education. In addition, these workshops should be held regularly in order to account for any changes in technology or the law that would alter how companies should comply with the proposed regulations.

Some helpful resources that could be used during this workshop regarding transparency, accountability and trust in the technology sector are the previously mentioned The Santa Clara Principles on Transparency and Accountability in Content Moderation, Transparency Reporting Toolkit on Content Takedown Reporting, Who Has Your Back report and the Ranking Digital Rights Corporate Accountability Index.

11: A new or existing regulator is intended to be cost neutral: on what basis should any cofunding contributions from industry be determined?

A regulator whose primary goal is to protect against harms to users should be an advocate for those users. There is a natural danger in such a regulator being funded by industry as such a direct financial relationship can lead to regulatory capture. As Bruce Schneier notes in his work analysing strategies to increase trust online, “If a government agency exists only because of the industry, then it is in its self-preservation interest to keep that industry flourishing.”[6]

[6] Bruce Schneier, *Liars & Outliers: Enabling the Trust that Society Needs to Thrive*, John Wiley & Sons.

But this touches on a broader point: the internet is *not* solely composed of the few giants of the (social media) "industry." A substantial proportion of the extremely wide category established within the scope of the regulator (i.e. entities that "allow users to share or discover user-generated content or interact with each other online") is not provided by major industry players. Small and medium-sized mobile applications, Web forums, group chat channels, emerging decentralised protocols like ActivityPub[7] and SOLID[8], and non-Web communication channels still occupy a sizeable proportion of the support of user interactions online. Many of these are operated by individuals as independent self-hosted projects, or supported by a distributed community with a loosely organized base.

These communities often provide, as a result of their size, the niche nature of their participants, or simply their cooperatively-run structure, a greater "duty of care" to their users as interpreted by this White Paper[9]. These communities can not be expected to provide financial support to the new regulator, but as importantly, may not be invited to the table in the negotiations between the regulator and the "industry" writ large, either because of their size, resources, lack of traditional organization, or location.

We believe that over and above issues of direct funding (which should be treated extremely cautiously), this risk of the regulator becoming over-receptive to the needs and approaches of internet giants, versus potential competitors and better solutions across the wider Web, is the most pressing danger for a new organization (or established UK regulator). This paper points to many challenges in the existing structure of the internet and digital communication: it should be alert to accidentally preserving that structure and prolonging the problems it seeks to address.

13: Should the regulator have the power to require a company based outside the UK and EEA to appoint a nominated representative in the UK or EEA in certain circumstances?

As its authors note, this paper is one of the first of what will be undoubtedly many proposals by different nations to create similar frameworks across the world. A common feature of those proposals will be to resolve the jurisdictional challenges by establishing a liable representative within their respective states.

While appealing as a simple solution, this proposal merely pushes the problem of jurisdiction into other domains which will have negative effects on online safety, trust, and human rights. The first is the creation of a burden on the companies: with the widespread adoption of similar frameworks, only a handful of major platforms will be able to afford a representative in *every* separate regulatory environment. This poses a limit on further competition — not just for new

[7] World Wide Web Consortium (2018), *ActivityPub*, https://www.w3.org/TR/activitypub/
[8] Tim Berners Lee et al, *Socially Linked Data* (SOLID), https://solid.mit.edu/
[9] Luke O'Neill, "Tired Of Nazis In Your Twitter Mentions? Try Mastodon", *Esquire UK*, https://www.esquire.com/uk/latest-news/a22798938/what-is-mastodon-twitter-platform/

competitors hoping to enter the UK market, but also for British companies seeking to expand outside the UK or EEA.

The second is only briefly highlighted in the paper, but will be a recurring challenge for a regulator whose obligations include defending free expression, privacy, the right of association and the right of access to information. If services are non-compliant, when and how should they be blocked or disrupted?

International human rights law provides very narrow circumstances under which free expression can be legitimately restricted by states, including online.[10] While the paper highlights edge cases where an argument may be conceivably made that such blocking is legitimate, the vast proportion of online services who are incapable of funding or organizing a nominated UK or EEA representative will not rise to this level. By establishing this as a requirement, the regulator will be either consigning itself to a UK online environment that is, under its own terms, mostly unlawful, or resigning itself to unlawfully blocking a disproportionate proportion of legitimate expression and information. We urge the government to consider forms of transparent and publicly accountable forms of international co-operation through established judicial processes, instead of compelling companies to establish legal structure in the United Kingdom, or risk violating international human rights law by blocking or disrupting foreign sites.

14. In addition to judicial review, should there be a statutory mechanism for companies to appeal against a decision of the regulator, as exists in relation to Ofcom under sections 192-196 of the Communications Act 2003?

Yes. There should be an appeal procedure similar to those outlined in the Communications Act that provides for reconsideration of the regulator's decision. The procedure should require review by someone other than the original decision-maker and yield a prompt decision. A time limit for the appellate decision, not more than 10 days, should be required for any original decision that is enforced during the pendency of the appeal.

Importantly, this procedure should be available not only to companies, but to any person whose content is flagged as "harmful content" by the regulator itself or otherwise subject to regulation. Indeed, a statutory mechanism for appeal, is especially important for persons who do not have the resources to engage in a protracted judicial proceeding. Thus, any appeal mechanism should be designed with minimum barriers to participation, and should not, for example, require an attorney.

18. What, if any, role should the regulator have in relation to education and awareness activity?

[10] E.g, United Nations Human Rights Council Res. 20/8, 16 July 2012, 20th Sess.,"The promotion, protection and enjoyment of human rights on the Internet" https://undocs.org/A/HRC/RES/20/8

One of the most compelling arguments made in the White Paper is the emphasis placed on user empowerment. We believe that education and awareness can play a part in this empowerment: but so should access to user-empowering technology, including third-party tools, independent self-hosted services, and innovative alternatives to the major platforms when the incentives of large, global and centralised commercial entities incorrectly match the desires or needs of their own users.

By supporting, advocating and educating British consumers on their rights with regard to interoperability, data portability, and using third-party filters and tools, the regulator will not only be empowering users, but also offering a useful market pressure on large platforms to more accurately represent the needs and interests of their users -- or risk having those users either move to other services, or take control back, and use their own algorithmic filters in preference to those provided by the platforms.

We appreciate the opportunity to submit these comments in connection with the Consultation regarding the Online Harms White Paper.

Sincerely,

Electronic Frontier Foundation:

Jillian York, Director for International of
Freedom of Expression

Danny O'Brien, International Director

www.eff.org

New America's Open Technology Institute:

Sharon Bradford Franklin, Director
Surveillance & Cybersecurity Policy

Spandana Singh, Policy Program Associate

www.newamerica.org/oti