Call for evidence response form

Your response – Additional terms of service duties

Questions 1 – 5: Terms of service and policy statements

For all respondents

Question 1: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements? Please submit evidence about what features make terms or policies clear and accessible.
Response: ToS in particular are usually designed to be clear to the business. Underlying policies refine how judgements are made. Because content decisions are done at speed, cost money and are also where possible automated, criteria behind these decisions can be very binary and lack nuance. For example, for Facebook nudity banned as “sexualised” content. However this was often determined by the presence of a female nipple. This led to removal of pictures of breast feeding which are in no way sexualised. Similarly, small amounts of copyrighted material are assumed to be copyright violations, where this may simply not be the case, for instance because it is incidentally included, or because it is used as part of review or commentary. In areas like removal of political material identified by slogans or logos this is especially problematic. Understanding the actual criteria applied as opposed to the policy or ToS position of what is allowed helps understand where the free expression friction is actually applied and where reviews may be needed. This should be done at the point that material is removed as well as in public documents. Understanding whether machines or humans made the judgement is also very important when deciding whether to ask for review. Barriers to reporting mistakes are high, and sometimes made worse by for instance demanding personal details to involve the person in a legal dispute, in the case of copyright. Where a decision has been made by a machine this is not appropriate. Policies need to be communicated where takedown occur in such a way that they can be clear that reviews do not come at any cost to the user and encourage reporting of mistakes.
Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 2: How do you think service providers can help users to understand whether action taken by the provider against content (including taking it down or restricting access to it) or action taken to ban or suspend a user would be justified under the terms of service? In your response to this question please consider and provide any evidence related to the level of detail provided in the terms of service themselves, whether services should provide user support materials to help users understand the terms of service and, if so, what kinds of user support materials they can or should provide.
Response: As above the actual reason and the mechanism (automated or human) needs to be explained when a decision is made. Access to further materials and support should be given at this point, for instance to explain what a copyright exception is and what might be allowed, or where commentary on an organisation does not violate terms of service banning its promotion. At all times, users should be encouraged to appeal bad decisions and given support to do so. This could include recommendations for independent advice, such as external websites or organisations that look at complaints. Examples of decisions and reasons should be given. It is critical that explanatory material is signposted when decisions are made as well as in ToS, etc. Larger companies should consider how they support external partners to help with difficult cases, such as helping people having difficulties to access external support. Content should not be de-prioritised in secret, or restricted to users without a clear reason given. This would enable users to challenge unfair decisions, and avoid situations where providers of content about topics uch as Addiction, substance abuse support content find themselves wrongfully or even unknowingly censored.
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Questions 9 – 15: Moderation

For all respondents

Question 9: Could improvements be made to content moderation to deliver more consistent enforcement of terms of service, without unduly restricting user activity? If so, what improvements could be made? **In your response to this question, please provide information relating to (a) –(c) where relevant.**
Response:

A key issue for content moderation is that business incentives do not align with performance and accuracy. For example

(a) provocative content is promoted as it attracts attention, but will often naturally fall close to content moderation boundaries as a result;

(b) accuracy requires both investment and time in human review.

(c) cultural familiarity with content is costly for instance requiring greater ranges of language familiarity, etc

A key change that could improve content moderation without impacting users would be to demand greater interoperability between platforms, so that users could receive and impart content from outside the current platforms. This could create market incentives for competing platforms that serve users better, and help such platforms or services to spread the load of moderation. Interoperability between Threads and Mastodon is a good example of how this can evolve. Users can choose between different moderated standards of servers, eg mastodon.social while receiving content from users on threads.net ; users do not have to accept Meta's approach to moderation while still being able to interact with their users.

(a) improvements in terms of user safety and user rights (e.g., freedom of expression), as well as any relevant considerations around potential costs or cost drivers;

Response: See above. Market drivers are as important if not more important than content regulation, given that moderation and user needs are currently seen as cost rather than investment, and improvements in moderation can result in reduced profits or revenues. If good moderation systems that serve users result in better user retention, then they will be seen as an investment, but user retention can only become a factor if interoperability is present and sufficiently robust so that users genuinely "leave" eg threads.net while not losing their networks or needing to rebuild them. Account portability as well as the ability to receive and send content across networks is critical to create this change.

Within systems, content removal requires incentives for accuracy as well as takedown. If content systems are seen largely as a matter of removal of content, and penalties are only imposed for unlawful or content violating ToS remaining present, then accuracy will tend to be sacrificed. Where external partners can file for takedowns, as with copyright and defamation, there is a lack of an incentive for accuracy. This can and does lead to bulk spurious takedown demands, while this could be simply prevented by imposing penalties on the actors that are filing thoughtless or unjustified claims. Ofcom needs to identify incentives it create within companies and any partner relationships to ensure that takedowns are accurate as well as encompassing the material the company seeks to remove.

(b) evidence of the effectiveness of existing moderation systems including any relevant examples of the accuracy, bias and or effectiveness of specific moderation processes;

Response:

(c) any other information.

Response:

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

our response – News publisher content, journalistic content and content of democratic importance

Questions 16 - 17: Identifying, defining, and categorising journalistic content, news publisher content and content of democratic importance

For all respondents

Question 16: What methods should service providers use to identify and define journalistic content and content of democratic importance, particularly at scale? In your response to this question, please provide information relating to (a) where relevant.
Response:
(a) how journalistic content and content of democratic importance can be described in the terms of service so that users can reasonably be expected to understand what content falls into these categories.
Response: There is no sensible, easy answer to this. Single issue campaign groups, Charities, non-party campaigners,bloggers, students, election candidates, elected officials and other commentators may be providing such content and cannot be identified in advance, or be expected to verify themselves in such a way. “At scale” attempts to verify one set of organisations over another will disempower marginalised voices over established commercial providers of news, which risks reinforcing power dis-balances in society at large; yet these dis-balances, including issues of diversity, equity and inclusion, are a large part of the agenda that the OSA seeks to address. Protection of journalistic content is one of the many problematic areas of the Act that are likely to cause significant dangers and problems as the Act is brought into force. Ofcom should be open to a rethink and request for changes in the legislative duties if they cannot find a way to implement it. At heart, the approach of the Act is for prior judgements to be made on the importance or harm of certain material, and for mechanisms, human or machine, to apply these value judgements. However, free expression law works the other way around: at the point that judgements need to be applied, considerations such as journalistic importance are considered and applied. To this extent, policies that aim at prior judgement, for instance prioritising certain outlets, are bound to be highly problematic. At a minimum, therefore, ToS’s need to ensure that any content about social issues, eelections, news-related or political content should be understood as potentially journalistic or of democratic importance.
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Questions 19 – 21: Complaints and appeal processes for journalistic content, news publisher content and content of democratic importance

For all respondents

Question 19: What complaint, counter-notice or other appeal processes should be in place for users to contest any action taken by service providers regarding journalistic content and content of democratic importance?

In your response to this question, please provide information relating to (a) and (b) where relevant.

Response: In all systems, takedown of material before a user can ask for review is deeply problematic. With political content, this is especially true. However, content in this area will cross over with disinformation and misinformation. Ofcom's approach to these questions needs to be cautious and ensure that measures protect free expression. As with comments above, interoperability, including "vertical" interoperability such as measures for users to use filtering and prioritisation engines of their own choice, or "horizontal" interoperability to receive and imart content from an entirely separate platform, are in our view likely prove more effective measures than relying on platforms to solve these issues from a single regulatory perspective. This is because users would be empowered to find better content and better moderation systems, and to remove themselves from the parts of content platforms that did not do their job effectively. The underlying points here were recognised in section 4.2 of Ofcom's 2023 paper on the topic.[1]

It is possible to try to see content in these areas as falling in four broad areas: (a) political content that can be fairly easily identified as from trusted sources; (b) political content with the same speech value that is not easily pre-identified and deserves the same protection; (c) misinformation spread by people who have fallen victim to rumours and exaggerations; and (d) deliberate poisoning of the political sphere by malign actors. Different strategies are needed for each, and there is a risk of designing policies that sacrifice particularly user driven speech as it is harder to verify.

Some problematic content that poses as political content is likely to have characteristics that are similar to other abusive content, such as spam and scams, including elements like automation. We would urge that any suggested action in this area focuses on strategies that are likely to be uncontroversial and do not pose dangers to users' speech before tackling harder problems. We would likewise caution against privileging certain actors' speech over others. Finally we would urge that Ofcom looks to work with competition authorities to leverage interoperability as a potential means to improve the overall information environment and incentives.

Outside of those areas where automation and other known features of spam-like behaviour may mean swift action can be taken, we would firmly believe that all users deserve pre notice that content may be removed, or deprioritised, before such action is taken. On filing a counter notice, the user's rights should then be protected until the issue is resolved. This protection should remain until at least a first full human review or appeal. Depending on the issue and likely harm, different content should either stay up or be taken down, depending on context. For example, potentially defamatory or copyright infringing content should stay up if a user desires, until a court has resolved the problem, if the user takes legal responsibility for it. Incentives for both

[1] https://www.ofcom.org.uk/__data/assets/pdf_file/0031/270589/Economics-Discussion-Paper-Mandated-interoperability-in-digital-markets.pdf

users and companies to allow truthful content and reasonable comment to stay up and / or be published must be sustained in any moderation system. Typically moderation systems lack such incentives, and reply on counter notices or appeals against takedowns, meaning that a small fraction of reasonable content that is wrongly removed is ever reinstated. Furthermore, reinstatement tends to be too late to be of much use. Thus systems should favour a 'stay up' policy where possible, in order to avoid harms to free expression from temporary removal.

(a) examples of effective redress mechanisms that you consider would be most suited to these content types

Response:

(b) briefings, investigations, transparency reports, media investigations and research papers that provide more evidence

Response:

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 20: What initiatives could service providers use to create and increase awareness about the process for users to complain and/or appeal content decisions and to minimise its' misuse?

In your response to this question, please provide information relating to (a) and (b) where relevant.

Response: Most users do not want restrictions imposed

Misuse of appeal content responses are likely to be from

(a) any known impacts of over-removal or erroneous removal of news publisher content, journalistic content or content of democratic importance

Response: We have seen many of these relating to copyright in particular, but also from youtube, Tik Tok and Facebooks' moderation systems. Youtube currently favours news outlets in reproduction for instance of war coverage and images that include potentially upsetting content. They operate demonetisation strategies which disfavour people exercising copyright exceptions for news reporting or review, as there is risk to contesting Youtube's decisions. Tik Tok algorithms will de-priotise or shadowban content that can be of democratic importance without any explanation given to a user. The impact of shadow bans is documented in a research paper by Horten, Monica, Algorithms Patrolling Content: Where's the Harm? (July 17, 2022). Available at SSRN: https://ssrn.com/abstract=3792097 or http://dx.doi.org/10.2139/ssrn.3792097

This is now unfortunately rife as a problem. Further issues come from lack of familiarity with language and culture. A level of discrimination against people from less privileged backgrounds emerges as a result To give an example educated 'in the know' users on Tik Tok will now use code words such as 'unalived' when discussing a news article rather than terms such as 'murder' or 'suicide' to avoid the automated moderation systems in place that might de-prioritise their content. The result is that systems that appear to be single tier with single content policies are increasingly nothing of the sort.

During elections, particularly local elections there have been cases of mass reporting content that individuals disagree with in an attempt to get posts censored. Moderators of groups on social media sometimes exclude candidates with opposing views to their own impacting on the fair and free functioning of elections.

To give a tangible example during the Conservative leadership election contest in 2022 Talktalk and Virgin media blocked Penny Mordaunt MPs website as being 'unsuitable'. Even when the site was later unblocked political damage and mockery of the candidate still occurs.

(b) briefings, investigations, transparency reports, media investigations and research papers regarding misuse of such speech protective provisions

Response:

Horten, Monica, Algorithms Patrolling Content: Where's the Harm? (July 17, 2022). Available at SSRN: https://ssrn.com/abstract=3792097 or http://dx.doi.org/10.2139/ssrn.3792097

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Questions 22 – 24: Other information for journalistic content, news publisher content and content of democratic importance

For all respondents

Question 24: What, if any, measures can online service providers put in place to ensure that protection of content of democratic importance applies in the same way to a wide diversity of political opinion?

In your response to this question, please provide information relating to (a) where relevant.

Response:

Approaches need to ensure that particular sources or publishers are not privileged over others. Freedom of expression is not a divisible concept or one where particular actors can be seen as safe while others are not. Rather, action needs to focus on bad actors, while treating all others in the same way.

(a) whether there are any additional measures/ safeguards that can be put in place during local or national elections

Response:

Political advertising should be a focus of regulatory attention but is beyond the scope of Ofcom's work. Within elections, transparency and accountability are paramount, as is preventing prioritisation and targeting to promote content, all of which can have very bad effects including voter suppression and fracturing the sense of local and national discourses.

Candidates need some protection against false reporting of their content resulting in wrongful censorship by platforms or from opposition groups. There is a danger that any wrongful moderation decision could be weaponised by opposition during an election period impacting on the integrity of the election. Platforms should ensure their community moderators are not

wrongfully blocking or censoring political candidates for example on a community Facebook group.
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – User empowerment duties

Questions 29 and 30: Features employed to enable greater control over content

For all respondents

Question 29: What features exist to enable adult users to have greater control over the type of content they encounter? In your response to this request, please provide information relating to (a) – (d) where relevant.
Response: As above, vertical interoperability measures to allow different engines for content sorting and moderation seems to be a key opportunity, along with horizontal interoperability to provide more extended methods of choosing different standards of content moderation. Within platforms, measures are likely to be less effective as they may compete with different priorities and business interests, unless competition and choice is allowed. Classification of content appears to be the understood direction of travel, but this is notoriously error prone and likely to restrict access, even by mere application of filters. If this is relied on without wider market measures, the results are likely to be sub optimal.
(a) features offered to users to reduce the likelihood of them encountering content they do not wish to see
Response:
(b) features offered to users to alert them to the presence of certain categories of content
Response:
(c) features offered to users to enable them to control their interactions with different types of users (e.g., non-verified)
Response:
(d) whether certain features are particularly valued or of use to users with protected characteristics, or by users likely to be affected by encountering relevant content
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – User identity verification duties

Question 34 and 35: User attitudes and demand for identity verification on user-to-user services

For all respondents

Question 34: What are user attitudes and demand for identity verification on user-to-user services? In your response to this request, please provide information relating to (a) – (d) where relevant.
Response: Generally we do not believe there is a demand for identity verification, and frequently the opposite is the case. In groups facing social prejudice and discrimination, anonymity is highly valued and often critical to be able to exercise free expression. Measures for identity verification may dissuade participation even where the technologies may be trustworthy, as users need to make their own judgements of risk. Where users do have to verify their identity for a service we see a clear demand for a high standard of privacy to protect against risks of online fraud and identity theft, and methods of verifying age that do not involve retention of personal data.
(a) whether they value verification being offered on a service
Response: This is more important where well known people or organisational accounts are seen. Here, there is value to ensure that some is who they say they are, as it improves trust. However, the utility for people to know that institutions and people with high profiles does not necessarily extend more generally. There is a good case for people having easy choices to do this, especially if it is easy and cost free. Some methods, eg with Mastodon verifying possession or control of relevant web pages, present some usefulness without causing intrusion or being seen as compulsory.
(b) whether verification influences user behaviour, such as whether they perceive identity verification to signify authenticity
Response: This depends very much on the person being interacted with. It is more important on decentralised platforms, such as Activity Pub and email. With email this tends to be achieved by the use of a relevant domain, for example. On larger platforms, there is an assumption that accounts are who they say they are, or that they will be removed. It also depends on what someone is trying to achieve in the interaction. There is not always an inherent need for someone to know very much about another user. The means to achieve sufficient understanding of who someone is can very greatly.

(c) attitudes towards non-verified, anonymous or pseudonymous users and the willingness to engage with them

Response: this is extremely context specific, but it should be understood that reputation online is built through behaviour and contact just as much as verification of the user. One example of this would be drug and addiction discussion groups where individuals might want to preserve anonymity, and posts will be judged by their individual merit of the content, and ratings on the usefulness of the user, and their post by other users.

(d) who you deem to be 'vulnerable' in terms of verifying their identity online – for example, whether this includes users unable to access or less likely to hold identification documentation, and those who may become vulnerable by displaying their identity to other users.

Response:

There are significant risks of exclusion if people are forced to verify their identity. This could disproportionately impact groups of society such as migrants, the elderly, or people on benefits. Vulnerabilities are well known, and include people active in politics and trades unionism worrying about their employer; people who have not disclosed their sexuality or gender identity to their relatives, friends or community, or who may suffer if their sexuality or gender identity is widely known; people avoiding former abusers or who are the victims of other kinds of crime; people with health conditions they wish to discuss on social media; and whistleblowers. Forcing people to be fully identified is a clear risk and would constitute a restriction of the free expression of many of the most vulnerable in our society.

Is this response confidential? (if yes, please specify which part(s) are confidential)

Response: