Call for evidence response form

Please complete this form in full and return to os-cfe@ofcom.org.uk

Title	Call for evidence: First phase of online safety regulation
Full name	Mark Johnson
Contact phone number	020 8075 8479
Representing (delete as appropriate)	Organisation
Organisation name	Big Brother Watch
Email address	mark.johnson@bigbrotherwatch.org.uk

Confidentiality

We ask for your contact details along with your response so that we can engage with you on this consultation. For further information about how Ofcom handles your personal information and your corresponding rights, see Ofcom's General Privacy Statement.

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? (delete as appropriate)	Nothing
Your response: Please indicate how much of your response you want to keep confidential. (delete as appropriate)	None
For confidential responses, can Ofcom publish a reference to the contents of your response? (delete as appropriate)	Yes

Your response

Please refer to the sub-questions or prompts in the annex to our call for evidence.

Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.	*Is this response confidential? – N* Big Brother Watch is a civil liberties and privacy campaigning organisation, fighting for a free future. We're determined to reclaim our privacy and defend freedoms at this time of enormous technological change. We're a fiercely independent, non-partisan and non-profit group who work to roll back the surveillance state and protect rights in parliament, the media or the courts if we have to. We publish unique investigations and pursue powerful public campaigns. We work relentlessly to inform, amplify and empower the public voice so we can collectively reclaim our privacy, defend our civil liberties and protect freedoms for the future.
Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?	*Is this response confidential? – N* The terms of service model, used by many large social media platforms to govern their sites, does not lend itself to accessibility or clarity when it comes to content moderation decisions. Whilst some platforms try to present their terms through more accessible "community guidelines", many do not or simply locate these pages where they are obscured from users' view. Often changes to platforms' rules are also published in places where users are unlikely to see them. When it comes to the permissibility of speech online, major internet intermediaries need digital constitutions that reflect the foundational values of the democracies they serve. This means

	content policies should reflect human rights principles and avoid limiting expression beyond the limitations of the law. These constitutions should clearly presented to users upon first access to the site, made accessible to users and should be referred to in all content moderation decisions. Currently, the terms of service model effectively gives most platforms absolute power and complete discretion as to their application of it. This needs to change. We believe that major internet platforms should adopt rule of law principles for enforcement. Government should be promoting rule enforcement that centres transparency of rules, foreseeability of their application, fairness of processes, the right to appeal, and equal and consistent application of the rules. In particular, when setting out their rules, platforms should make the text easy to understand. Rules should be clearly defined and refrain from being subjective. Users should be actively notified by the platform as to any rule changes. By ensuring that rule of law principles are embedded in platforms' processes, in a way which is clear to users, fundamental rights can be protected online.
Question 7: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?	*Is this response confidential? – N* User's right to redress is a key aspect of platforms' content moderation processes. Whilst it is welcome that the Online Safety Bill compels online intermediaries to offer an appeals process to users who have had content restricted, in practice this is already the case on most major social media platforms. However, these processes are often ineffectual, automated, lack clear process and content is rarely assessed in the full

	context in which it was posted. The Government and Ofcom should create minimum standards for platforms appeals processes. The Santa Clara Principles (2021), drafted by human rights organisations and academics establish some basic principles for centralised content moderation systems, to ensure they are compliant with human rights standards. The principles state that user notice to those who have contravened a platform's rules should include the following: • URL, content excerpt, and/or other information sufficient to allow identification of the content actioned. • The specific clause of the guidelines that the content was found to violate. • How the content was detected and removed (flagged by other users, trusted flaggers, automated detection, or external legal or other complaints). • Specific information about the involvement of a state actor in flagging or ordering actioning. Content flagged by state actors should be identified as such, and the specific state actor identified, unless prohibited by law. Where the content is alleged to be in violation of local law, as opposed to the company's rules or policies, the users should be informed of the relevant provision of local law. The Santa Clara Principles also state that appeals processes should incorporate the following: • A process that is clear and easily accessible to users, with

	details of the time-line provided to those using them, and the ability to track their progress. • Human review by a person or panel of persons who were not involved in the initial decision. • The person or panel of persons participating in the review being familiar with the language and cultural context of content relevant to the appeal. • An opportunity for users to present additional information in support of their appeal that will be considered in the review. • Notification of the results of the review, and a statement of the reasoning sufficient to allow the user to understand the decision. Only through intermediaries following due process and applying a rules-based approach to content moderation can users rights be fully respected online.
Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?	*Is this response confidential? – N* As mentioned in response to question 5, in order to guarantee the right to freedom of speech, platforms' content moderation policies should reflect human rights principles and not curtail expression beyond the limitations of the law. Ofcom must remain mindful that obliging foreign intermediaries to uphold their frequently changing corporate content restrictions, that are more restrictive than domestic law when it comes to the permissibility of speech, would risk breaching the body's obligation to uphold freedom of expression under Article 10 of the ECHR. We believe that platforms should work

	from digital constitutions, ensuring that due process and a rule of law approach is bedded into the site's systems and processes. As well as protecting the right to freedom of expression, this approach also creates clarity for users and allows for greater transparency in terms of the application of rules. Platforms should also consider the types of moderation techniques they use carefully. An emphasis on human-lead moderation is key to protecting the right to free expression online. Automated systems often fail to detect context or nuance and can result in over-removal. Such systems will also suffer inherent biases based on their design which could lead to disproportionately negative outcomes for users from minority groups. In this context, it is also crucial to note that Article 22 of the UK GDPR states that: "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".
Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	*Is this response confidential? – N* The largest online intermediaries frequently terminate or suspend the accounts of users without warning, notice or explanation. When Impress regulated left-wing media outlet Novara Media had their entire channel deleted in October 2021, they said this had been done "without warning or explanation". This is an entirely inappropriate way for social media companies to conduct content moderation. A departure from the terms of service model and an approach based on rule of law principles would ensure not only consistent application of rules but

	adequate notice when rules on a given site have been breached.
Question 18: Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?	*Is this response confidential? – N* The Online Safety Bill includes clauses which sanction the state to issue notices to online intermediaries, compelling them to use "accredited technology" to search for certain kinds of illegal material on both public and private channels, without suspicion. This clause has the potential to seriously threaten the rights to privacy and freedom of expression. There are important technical issues to consider when imposing a "duty of care" on companies' private messaging channels. Some companies offer structural privacy to their services – for example, the end-to-end encryption offered by instant messaging/VoIP apps WhatsApp and Signal. It is concerning that the Government's intentions appear to deliberately make privately designed channels of this kind incompatible with platforms' obligations set out in the Bill. It is vital that terrorism and CSEA content are removed from the internet. However, tackling such content does not require encrypted channels to be compromised, sacrificing the security, safety and privacy of billions of people. Given that private messaging services are within the scope of the legislation, the provision above does imply that certain types of technology could be used to break, erode or undermine the privacy and security provided to messaging services by end-to-end encryption. This could involve the use of a technique known as client-side

	scanning, which would create vulnerabilities within messaging services for criminals to exploit or could open the door to a greater level of surveillance through use of this technology. It is not unreasonable to expect that such technology would be escalated in time, put to use in other areas and result in increased surveillance of individuals' private messages. The Online Safety Bill and regulatory framework would not comply with international human rights standards if it forced intermediaries to use technology which would undermine end to end encryption, such as client-side scanning.
Question 22: What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?	*Is this response confidential? – N* Conventional social media sites should not mandate age verification in order to gain access to their sites. Such systems are likely to involve verification using identifying documents like driving licenses or passports. This undermines citizens' right to privacy and the ability to navigate the internet anonymously, which is crucial to the safety and security of many internet users. Online anonymity is vitally important to journalists, human rights activists and whistleblowers in the UK and all over the world. Even tacit attempts to undermine online anonymity here in the UK would set a terrible precedent for authoritarian regimes to follow and would be damaging to human rights globally. Other groups that may rely on anonymity in order to use the internet safely, both at home and abroad, include LGBT people, survivors of domestic abuse, employees who may not wish to disclose their political views to employers and staff from organisations who want to make public disclosures about negligence or

	bullying. Such a measure would also mean that internet users would have to volunteer even more personal information to the platforms themselves, which could be stored in large databases. Further, many people across the UK do not own a form of ID and would directly suffer from digital exclusion. The United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has said that states should not restrict anonymity because of the important role it plays in a democratic society, but instead ensure that legislation "recognize[s] that individuals are free to protect the privacy of their digital communications by using (...) technology and tools that allow anonymity online".
Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	*Is this response confidential? – N* Platform transparency is crucial to user empowerment online. At present, the relationship between intermediaries and users is too asymmetrical. Giving users a greater degree of understanding of the processes of online intermediaries would empower individuals to exercise greater agency over their choices when navigating the web. Platforms should publish regular transparency reports with comprehensive thematic content moderation statistics. These reports should set out the number of account deletions, account suspensions, pieces of content taken down, pieces of content labelled, appeals made and content moderation decisions overturned from the reporting period. Platforms should explain in the clearest terms to users how their algorithms work and give users the option to see content in alternative

	forms, for example chronological ranking rather than suggested ranking based on engagement. Platforms should also grant researchers and members of civil society organisations as great a degree of access to the design of such processes as is possible. States are increasingly utilising the power of centralised social media companies by forcing platforms to engage in extra-judicial censorship on their behalf. In order to help protect users' fundamental rights, platforms should publish records of any state interventions in content moderation decisions. States should not extra-judicially censor or force the removal of content outside of formal legal frameworks. Where the state has intervened on an individual piece of content, the user who's online expression has been affected should be informed that such intervention has occurred.
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	*Is this response confidential? – N* We have made clear our view that the state outsourcing online policing to social media companies is a flawed approach. Social media companies are ill-judged to make determinations on illegality and this approach undermines the due process of the UK's legal systems. The Online Safety Bill places additional obligations on social media companies to police speech on their sites. A better approach would be create means of facilitating a smoother relationship between platforms and law enforcement agencies. Outsourcing online policing to the platforms themselves also threatens to prevent victims of online crime from finding a path to justice. If the only obligations on platforms are to take down the offending material in question, then they will do no more

	than is necessary. This could mean deleting evidence in the process and leaving perpetrators at large. Policy-makers should explore new means of connecting users who feel that they have been victims of crime online with law enforcement agencies. In order that users have a sufficient pathway to justice, law enforcement agencies should be properly resourced and be specially trained in the complex nature of online crime.

Please complete this form in full and return to OS-CFE@ofcom.org.uk