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:

One-third of all internet users are children,[1] and as such it is important that tech companies ensure their terms of service are accessible for children of all age ranges and regardless of who they are, what their background is, or where they are from. Published terms cannot be 'one size fits all' and having age-appropriate, concise, and well-formatted published terms are central to giving users agency and knowledge regarding the agreement they are entering into when they use a service.

5Rights research[2] has shown that many services popular among children and young people set out their terms of service in highly legalistic documents, sometimes over 11,500 words in length, with a 'readability' score requiring a university education.

To ensure services providers' published terms are age-appropriate to children and meet their duties in the Online Safety Act, organisations must refer to standards informed by consultation with children themselves, as well as technical experts, such as the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA) on Standard P2089 for an Age-Appropriate Design Framework.[3] Further, services must meet standards of the Age Appropriate Design Code[4] to provide published terms, policies, and community standards in concise, prominent formats with language suitable to the varying ages and capacities of child users, and ensure they are upheld.

Service providers should consider the following when creating or updating terms or policies:

Language

The language used in published terms is often inaccessible to children and is complex to read, time-consuming, and unclear. To aid children's comprehension and understanding, published terms must:

- **Use accessible language** and concepts appropriate to the age or age-range of a user, considering the evolving capacities of children. Published terms must avoid using jargon or complex, legalistic language.
- **Spell-out key terms and make them prominent**, ensuring they are understandable to the youngest users permitted on the service.

[1] Livingstone, S., Carr, J. & Byrne, J. (2015) One in Three: Internet Governance and Children's Rights. *Global Commission on Internet Governance.* Paper no. 22

[2] 5Rights Foundation (2021) *Tick to Agree: Age appropriate presentation of published terms*

[3] IEEE SA (2021) IEEE Standard 2089 for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children

[4] Information Commissioner's Office (2021) Age appropriate design: a code of practice for online services

- **Adopt language that makes clear what children are agreeing to**, ensuring that published terms aid comprehension and have been understood by them.[5]

Length

Published terms are often long – so much so they are often difficult to digest for even the average adult user. Indeed, research by Ofcom found that the length of terms on services used frequently by children, including TikTok, Snapchat and Twitch, would take roughly 20 minutes or more to read.[6] 5Rights research[7] found that children were consistent in saying that they would be more likely to understand published terms if they could find the most important parts.

In order to ensure published terms can be read by children, published terms must:

- **Be concise**, to the point and be short in word count. Shorter published terms must still include the most important information.
- **Divide terms into clear sections**, and/or be made available in bite-sized pieces.[8]

Format

Published terms are often presented in a single format, typically large blocks of text, which are difficult for children to read. 5Rights [9] research found that children preferred other formats that made published terms easier to understand, including images, videos, animations, audio and graphics. The Information Commissioner Office's (ICO) Age Appropriate Design Code[10] recommends that services use formats that will "attract and interest children", such as diagrams, cartoons, gamified or interactive content.

Further, published terms typically adopt a 'one-size-fits-all' approach, which fails to consider children's evolving capacities as they develop, as well as varying abilities among children. A single set of terms is unlikely to serve both older and younger children simultaneously.[11]

To ensure formatting is age-appropriate, published terms must:

- **Present terms in multiple formats for different ages**, utilising non-textual formatting where appropriate to help with children's comprehension.
- **Consult with children** to determine the most appropriate methods for helping children to understand terms and policies.[12]

Navigability

Published terms are not always prominent and difficult to find. For example, the livestreaming platform Twitch comprises several documents encompassing its published terms[13] including its terms of service, community guidelines, and privacy notices. 5Rights research found children find it

[5] IEEE SA, Std. 2089, §12.3(3)
[6] Ofcom (2023) Regulating Video-Sharing Platforms: What we've learnt about VSPs' user policies, pp. 11-17
[7] 5Rights, *Tick to Agree*, pp. 12-14
[8] IEEE SA, Std. 2089, §12.3(a)(1)
[9] 5Rights, *Tick to Agree*, pp. 14-15
[10] Information Commissioner's Office (2021) Age appropriate design: a code of practice for online services, Standard 4, Transparency
[11] Information Commissioner's Office (2022) *The Children's Code design guidance*
[12] Ibid, pp. 8-13
[13] Twitch, Twitch Policies and Guidelines

difficult to know what they are agreeing to when oversimplified text is summarised at the bottom of terms, but found it to be an inconvenience when covering a large portion of the screen.[14]

To ensure published terms are easy to navigate by children, published terms must:

- **Be prominent and easy to find**, ensuring they are easily locatable by children.
- **Make key terms and definitions searchable** to allow children to find the information they need or want to know.

Timing

The timing of published terms is usually at point of download, registration, or first use, meaning users have little or no opportunity to review them after using the service. Research from the Competition and Markets Authority[15] found that less than 5% of users that registered for Facebook engaged with privacy and ad preferencing settings in the first 28-days. For children, timing is critical to their understanding of published terms, with children saying accepting terms was "inevitable" to use the service, even if they had not read them.[16]

Further, published terms are generally difficult to locate after they have been agreed to and services often responsibilise users – including children – to refer and review published terms where there are updates to its policies.[17]

To ensure the timings of published terms are appropriate and aid in children's comprehension of agreements, published terms must:

- **Be presented at multiple points in the user journey**, including in regular intervals and at crucial moments when consent is sought. For lower-risk services, a basic overview of terms at the initial point of download with an option to 'edit details later' may be appropriate. This allows children to review published terms at a more convenient time for them.
- **Signpost children back to terms or policies** ensuring they are easily locatable – particularly if they are updated or changed.
- **Align with the Age Appropriate Design Code**,[18] ensuring policies are communicated during key moments in the user journey (e.g. onboarding, adding or removing features, reviewing or changing settings, product updates, or deleting user accounts/data) particularly at points when features that use children's data, such as geolocation, are active, or when it is used for personalisation and/or is shared with other users or third parties.

Accessibility

Published terms are primarily geared towards English-speaking users making them inaccessible to users, including children, where English is not their first language.

Further, published terms are rarely geared generally towards children, relying on the incorrect presumption that 'engaged' adults are present in a child's digital life. Published terms on services

[14] 5Rights, *Tick to Agree*, pp. 15-16
[15] Competition and Markets Authority (2020) Online platforms and digital advertising, p. 175
[16] Written evidence submitted by the Horizon Digital Economy Research Institute, The University of Nottingham (SMH0131), 5(22)
[17] 5Rights, *Tick to Agree*, pp. 16-18
[18] Information Commissioner's Office (2021) Find the best moments to engage children with privacy information

that children are likely to access, such as the Pokémon forum Bulbagarden,[19] are written exclusively for parents and fail to tailor information directly to children.

Published terms do not always have consideration for children with specific characteristics, lacking formats that certain children want to aid their understanding, such as diagrams, pictures, 'read aloud' features, and bright colours.[20] Published terms must have regard to children's evolving capacities as per the UN Convention on the Rights of the Child[21] and its General comment No. 25 in relation to the digital environment,[22] and for the "universal accessibility" of all children.[23]

In order to ensure published terms are accessible to all children, regardless of background or capabilities, published terms must:

- **Consider the diverse needs of young people**, providing terms in multiple languages and catering for children with accessibility requirements. Service providers should refer to technical standards in IEEE Std. 2089[24] that consider children's characteristics (age, gender, ethnicity), context (urban, rural, geography, language) and circumstances (device-sharing, cost of connectivity, children without adults), and in accordance with the latest Web Content Accessibility Guidelines (WCAG).[25]
- **Be accessible to all children without need for an adult**, making use of bold text, graphics, and icons where appropriate, consulting with diverse groups of children on formatting and design decisions.

Ensuring meaningful consent

Consent must be sought and obtained, not assumed, and this is crucial for aiding children's comprehension of what terms or policies they are agreeing to. Guidance from the ICO recognises consent must be given by a "clear affirmative act establishing a freely given, specific, informed, and unambiguous indication."[26] 'Tick box' or 'unread' consent is **not** appropriate for child users.[27]

Children cannot fully understand a service's terms or policies unless they are given meaningful choices about what parts they wish to interact with. There is often no option to only accept certain terms or conditions, or refuse non-essential ones. Rather, users – including children – are required to blanket 'agree' or 'disagree' in order to access a service. The most prominent example of this are cookie banners, such as those used on the gaming service Steam,[28] which only offer the option to "accept all" or" reject all" without explaining the implications of either decision. This is despite concerns children have about their data.[29] Consultations held by 5Rights[30] with children find they are frustrated when they are unable to agree to individual policies in published terms.

---

[19] Bulbapedia, Privacy policy
[20] 5Rights, *Tick to Agree*, p.19
[21] United Nations (1989) United Nations Convention on the Rights of the Child, Article 5
[22] United Nations Committee on the Rights of the Child (2021) General Comment no. 25 on children's rights relation to the digital environment, §4
[23] Ibid, §9
[24] IEEE SA, Std. 2089, §12.3(4)
[25] World Wide Web Consortium (2023) Intro to Understanding WCAG
[26] Information Commissioner's Office (2021) What is valid consent?
[27] 5Rights, *Tick to Agree*, pp. 22-28
[28] Steam
[29] 5Rights, *Tick to Agree*, pp. 22-28
[30] Ibid

The ability for children to give meaningful consent and make meaningful choices is crucial to the realisation of their rights in the digital world.

In order to ensure meaningful consent has been obtained, published terms must:

- **Give children the ability to give their consent or allow them to withdraw it** from parts of services they do not use, using features such as nudges, to empower them in making these decisions and understand what the implications this would have.[31]
- **Ensure children have meaningful choices** and have the option to refuse certain policies without impeding their access to the rest of the service.

Upholding published terms

In order for children to understand terms and policies, service providers **must** clearly uphold them; the formatting and presentation of published terms are only part of the wider picture that helps children to comprehend and understand what they sign up to. If service providers do not follow their own published terms, it is difficult for other users – including children and parents – to know and understand them. Research in 2020 by the Wall Street Journal into Facebook shows that content promoting violence and/or spreading misinformation remained on the site after it was flagged, despite it violating Facebook's own guidelines.[32] 5Rights research[33] illustrates how children's exposure to harmful content is common, stemming from recommender systems and algorithms that pushes this to them.

The Age Appropriate Design Code[34] states that services "should provide information that is accurate and does not promise protections or standards that are not routinely upheld." Published terms should adhere to industry codes or guidance on children's safety, and terms must be fully aligned with reporting mechanisms.

As examples of good practice, the social media service Yubo has an industry-leading Safety Hub, presented in an age-appropriate way, with guides and resources for how to behave on the service.[35] Twitter's Data Dash presents privacy policy information in a gamified format, informing users how their data will be collected, used, and how it can be protected.[36]

5Rights research[37] contains more details on the techniques referred above, as well as a "tick to include" checklist which service providers can deploy to make their terms suitable to all children.

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

Response: No

[31] See: Information Commissioner's Office (2021) Age appropriate design: a code of practice for online services, Standard 13, Nudge techniques

[32] Horwitz, J. (2020) Facebook Has Made Lots Of New Rules This Year. It Doesn't Always Enforce Them. *The Wall Street Journal*

[33] 5Rights Foundation (2021) *Pathways: How digital design puts children at risk*, pp. 66-86

[34] Information Commissioner's Office (2021) Age appropriate design: a code of practice for online services, Standard 4, Transparency

[35] Yubo Safety Hub

[36] Twitter Data Dash

[37] See: 5Rights, *Tick to Agree*, pp. 36-37

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:

Any information provided to children by a service, including information regarding reports they have made, must be age-appropriate, designed so that they are comprehensible, an appropriate length, clearly presented, and understandable to all children – no matter their age of where they come from. Children must be given information and supported throughout to understand how complaints are being actioned – children often say they are deterred from reporting as they do not believe reporting systems are effective.

Research by the VOICE Project[38] found that approximately 3-to-4 to children were aware of when to use support tools like reporting, but views on its effectiveness were mixed. Research commissioned by Ofcom[39] found that children feel reporting rarely leads to take-downs and that if reports were met with a quick response, they would feel more confident in using these systems. In the long-term, it also found that children don't understand what happens after they report and that there was no explanation for actions taken.

If a regulated service provider takes action that impacts a child, it must communicate this in a way that is age appropriate, provide sufficient information so the child understands what has happened and why, and explains their right to appeal, and the process for doing so. Further, where a child who is vulnerable has had action taken against them, service providers should also ensure they are signposted or directed to appropriate safeguarding resources.

When parents want to report content on behalf of a child, or regarding content which could be harmful to them, services must clearly communicate the action, if any, they have taken and the rationale for doing so. At 5Rights, we have heard from parents that when they report harmful content (e.g. 'challenge' videos), they are told it does not violate the service's terms – e.g. community guidelines – and are provided with no further explanation. If services determine content does not violate its published terms, it must set out clearly in its response to parents or responsible adults who may have made a report on behalf a child, why this is the case, and what further action they can take – including options to redress or appeal these decisions.

These provisions must also extend to non-registered users of the service and allow them to submit a report and be provided an explanation where content has not been taken down. Furthermore, one parent also recounted that she was still able to see content after it was reported, despite having indicated to the service that it was harmful.

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

Response: No

[38] The VOICE Project (ECPAT, Eurochild, Terre des Hommes Netherlands & Down to Zero) (2024) *Speaking up for change: Children and caregivers voices for safer online experiences*, pp. 62-63

[39] Ofcom & YouGov (2024) Children's Attitudes To Reporting Content Online

For providers of online services

Question 3: How do you ensure users understand the provisions in your terms of service about taking down content, restricting access to content, or suspending or banning a user from accessing the service and the actions you might take in response to violations of those terms of service? In your response to this question, please provide information relating to (a) – (d) where relevant.
Response:
(a) how you ensure your terms of service enable users to understand both what is and is not allowed on your service, and how you will respond to user violations of these rules;
Response:
(b) any relevant considerations about the risk of bad actors taking advantage of transparency around your terms of service and how they are enforced;
Response:
(c) details about any user support materials or functionalities you provide to assist users to better understand or navigate your terms of service or related products;
Response:
(d) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 4: Please describe the processes you have in place to measure user engagement with and comprehension of your terms of service and how you make improvements when required. In your response to this request, please provide information relating to (a) – (f) where relevant.
Response:
(a) how you measure user engagement with/comprehension of your terms of service and the metrics you collect;
Response:
(b) any behavioural research you undertake to better understand engagement with and/or comprehension of your terms of service (including any research into reasons why users do not engage with terms of service);
Response:
(c) any measures you have taken to improve engagement with and/or comprehension of your terms of service, including (but not limited to) how the findings of any behavioural research influenced these measures and/or any design changes (e.g. prompts to remind users to read the

terms of the service, changes to the structure of the terms of service or changes to how users access the terms of service etc.);
Response:
(d) costs of these processes (including the design, implementation and continued use of these processes or updated versions of these processes);
Response:
(e) how you evaluate the effectiveness of measures designed to improve engagement with and/or comprehension of your terms of service;
Response:
(f) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 5: Please describe any evidence you have about the effectiveness of using different types of mechanisms to promote compliance with terms of service or change user behaviour in the event of a violation, or potential violation, of terms of service. In your response to this request, please provide information relating to (a) – (d) where relevant.
Response:
(a) any evidence about the effectiveness of enforcement measures such as taking down content, restricting access to content, or suspending or banning user accounts in relation to encouraging users to comply with specific aspects of terms of service in the future
Response:
(b) any evidence about how effective non-enforcement mechanisms are at reducing violations of the terms of service or repeated violations, including the type of non-enforcement mechanism and how it is implemented (e.g. prompts for users to consider the appropriateness of their content before posting it to the service (with or without links to specific provisions within the terms of service), or prompts for users to review certain provisions within the terms of service when their content is found to violate these provisions)
Response:
(c) any information and/or evidence on the costs of designing and implementing different types of enforcement or non-enforcement mechanisms (including costs of the research behind the design, implementation and continued assessment/study of these mechanisms)
Response:
(d) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Questions 6 – 8: Reporting and complaints processes

For all respondents

Question 6: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?

In your response to this question, please provide evidence about what features make user reporting and complaints systems effective.

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

Response:

Inadequate reporting and complaints mechanisms are central to the widespread lack of trust children have in online services, with children frequently unable to report other users for violations of terms and community guidelines and frustrated at the lack of transparency around the complaints process. Indeed, recent research commissioned by Ofcom found that reporting mechanisms were widely understood by children, but they did not think they were effective, took too long, and did not encourage them to engage with these systems.[40] Research by UCL found that 30% of young people who had experienced unwanted sexual attention or had an image shared without their consent did not report this to the service as they did not think it would work.[41]

Similarly, parents also do not feel that reporting systems are adequate enough to allow adults to report on behalf of children. Testimony from bereaved families has reflected how difficult it is for them to navigate reporting mechanisms and find a human in the system when the worst has happened.[42]

To provide adequate reporting tools for children and their parents, services should:

- **Have prominent and accessible tools that are age-appropriate** and tailored to different age ranges, being easy to find for both children and parents. Where children or their parents are moved to report, this must be available at the point of worry – on the front-page or homepage of a service – to both registered and non-registered users.
- **Reflect children's lived experiences** by giving children the ability to explain why they are reporting, drawing on from social cues that they recognise, and ensuring systems use child-friendly language.
- **Signpost children** to reporting, complaints, and redress mechanisms during the sign-up process (e.g. highlight them and provide instructions on how to use them) as well as throughout the user experience.
- **Provide access to expert advice** where children and their parents have been moved to report, complain, or seek redress, to help them understand their rights and support their decision-making.

[40] Family Kids & Youth (2024) Understanding Pathways to Online Violent Content Among Children

[41] Ringrose, J., Regehr, K. & Milne, B. (2021) Understanding and Combatting Youth Experiences of Image-based Sexual Harassment and Abuse. *UCL, School of Sexuality Education, University of Kent, ACSL*, p. 50

[42] Bereaved Families for Online Safety (2024) Submission for Ofcom's illegal harms consultation. See also: 5Rights Foundation (2024) Ofcom: Protecting people from illegal harms online, pp. 23-25

- **Provide expected response times** that are proportionate to the seriousness of the report made, with reports made by children, or on their behalf, expedited and prioritised as a matter of urgency, ensuring children are given heightened protections.
- **Have a human involved early-on in the process** where a report is made in relation to children's safety.

When designing reporting tools, service providers should:

- **Refer to existing technical standards for age-appropriate design**, such as IEEE Std. 2089,[43] to ensure that systems use accessible language, provide multiple formats, consider the diversity and evolving capacities of children.
- **Consider the barriers to children making reports** and give them options that encourages them to make use of such mechanisms – e.g. allowing for anonymous reporting.
- **Guarantee data protection and privacy** when using reporting systems, with processing of personal data done in line with GDPR and guidance under the ICO, as well as in accordance with statutory codes like the Age Appropriate Design Code where this applies to the processing of children's data.
- **Consider using *Trusted Flagger* programmes** to support children and their parents to raise issues about the service.

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

Response: No

For providers of online services

Question 7: Can you provide any evidence or information about the best practices for effective reporting and/or complaints mechanisms, and how these processes are designed and maintained?

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

Response:

(a) how users report harmful content on your service(s) (including the mechanisms' location and prominence for users, and any screenshots you can provide);

Response:

(b) whether there are separate or different reporting or complaints mechanisms or processes for different types of content and/or for different types of users, including children;

Response:

(c) how users appeal against content takedowns, content restrictions or account suspensions or bans;

Response:

[43] IEEE SA, Std. 2089, §12.3

(d) what type of content or conduct users and non-users may make a complaint about / report, including any specific lists or categories;
Response:
(e) whether users need to create accounts to access reporting and complaints mechanisms (if there are multiple mechanisms, please provide information for each mechanism);
Response:
(f) whether reporting and complaints mechanisms are effective, in terms of: **(i) enabling users to easily report content they consider to be potentially the types of content specified in the relevant terms of service, and how to determine effectiveness;**
Response:
(ii) enabling, supporting or improving the accuracy of user reporting in relation to identifying the types of content specified in the relevant terms of service, and how to determine effectiveness;
Response:
(iii) enabling, supporting or improving the provider's ability to detect and take timely enforcement action against content or users as specified in the relevant terms of service, and how to determine effectiveness;
Response:
(g) whether there are any reporting or complaints mechanisms you consider to be less effective in terms of identifying certain types of content and how you determine this;
Response:
(h) the use of trusted flaggers (and if reports from trusted flaggers should be prioritised over reports or complaints from users);
Response:
(i) the cost involved in designing and maintaining reporting and/or complaints mechanisms, including any relevant issues, difficulties or considerations relating to scalability; and
Response:
(j) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 8: What actions do or should services take in response to reports or complaints about content that is potentially prohibited or accounts engaging in potentially prohibited activity? In your response to this question, please include information relating to (a) – (g) where relevant.
Response:

(a) what proportion of reports are reviewed, and what proportion result in action taken including; **(i) any potential variation in the number and actionability (i.e., the proportion that result in a takedown or other action) of reports or complaints in relation to different provisions within your terms of service;**
Response:
(ii) any differences for cases involving multiple reports/complaints about a single piece of content or user;
Response:
(iii) the costs associated with reviewing reports;
Response:
(b) whether any reports or complaints are expedited or directed to specialist teams, including: **(i) the criteria for this;**
Response:
(ii) the cost involved in facilitating this;
Response:
(c) the extent to which relevant individuals (content creators, users, and non-registered or logged-out users) are informed about the progress of their report or complaint, including: **(i) if they are not, the reasons why;**
Response:
(ii) if they are, what is included when users are informed about the progress of their report (e.g. receipt of the report, the progress of the report through the service's review process, and/or the outcome of the report);
Response:
(iii) the technical mechanisms/process to inform any relevant individuals about the progress of their report (e.g., whether non-registered users are provided an opportunity to provide an email address);
Response:
(iv) any differences in responses to different types of reports (e.g., reports about content or an account a user believes violates the terms of service, about the provider not operating in line with its terms of service, or about the accessibility, clarity or comprehensibility of those terms of service);
Response:
(v) the costs associated with responding to reports;
Response:
(d) what happens to the content while it is being assessed/processed (e.g., if and how it may still be found or viewed by other users);

Response:
(e) any internal or external timeframes or key performance indicators (KPIs) for reviewing and/or acting on reports or complaints;
Response:
(f) any user support materials that are used or should be used to support users understand the service's responses to reports, or how users can appeal moderation decisions about their content or accounts, or about decisions taken in response to reports they have submitted about other users' content or accounts;
Response:
(g) any other information.
Response:
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: Moderation systems must meet the needs and expectations of children and parents who act on their behalf, upholding terms and allowing clear means to redress or challenge decisions where adverse outcomes are reached. Building trust in these systems is key to ensuring children have the confidence to report content. This is essential given the importance Ofcom has placed on user reports as part of a 'core input' in the risk assessment for both illegal harms and harms to children. When designing or enforcing moderation systems, providers should refer to the activities and tasks in IEEE Std. 2089[44] which sets out standards for the age-appropriate application of these systems. It requires that services: • **Provide clear means to redress** that are prominent and accessible, provide children and parents advice where needed, apply penalties fairly and consistently, offer opportunities to appeal and escalate unresolved appeals, provide reasonable response times and offer the right to correct digital footprint and/or termination rights. • **Uphold terms clear and unambiguously** by informing children of action taken during redress processes, obtaining valid consent for upgrades and amendments to the service,

[44] IEEE SA, Std. 2089, §11.3

publishing corporate policies and reviewing matters that arise through moderation systems. • **Enforce fair terms** by not introducing (or re-introducing) unfair terms and recording obstacles to moderation and redress systems.
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

For providers of online services

Question 10: Please describe circumstances where you have taken or would take enforcement action against content or users outside of what is set out publicly in your terms of service and the reasons for taking this action. In your response to this question, please provide information relating to (a) – (e) where relevant.
Response:
(a) the types of action taken, and frequency of these actions (including per type of action);
Response:
(b) how relevant content or users were or would be brought to your attention;
Response:
(c) any policies, approaches or processes you have used or would use to guide moderation decisions in these cases;
Response:
(d) whether new policies are or would be written in response to these cases, and if so: **(i) whether and when these new policies are written before enforcement action is taken or after;**
Response:
(ii) when and how these new policies would be added to or included in your publicly available terms of service;
Response:
(e) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 11: If you are made aware of content or an account that potentially violates your terms of service, please describe any relevant circumstances which might not result in enforcement action, immediately or at all.

In your response to this question, please provide describe (with examples) any relevant circumstances relating to (a) – (e).
Response:
(a) circumstances that relate to issues or challenges within your content moderation system (e.g. moderator error, language or local knowledge gaps, content is no longer available (e.g. livestream), nuance/context of content means it is found non-violative, further investigation needs to be done before action can be taken);
Response:
(b) circumstances that relate to issues or challenges within your terms of service and/or associated policies (e.g. new iterations of a harm falls outside the scope of internal moderation policies, individual piece of content is only of concern at scale (but itself does not violate policies);
Response:
(c) circumstances that relate to competing priorities (e.g., freedom of expression, public interest concerns);
Response:
(d) circumstances that would be understood by a user who has read the terms of service and why or why not, (e.g., the terms of service sets out exception for not removing violating content (e.g. news content), or transparency is not provided to avoid empowering bad actors);
Response:
(e) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 12: What automated systems do you have in place to enforce terms of service provisions about taking down or restricting access to content or suspending or banning accounts? **In your response to this question, please provide information relating to (a) – (d).**
Response:
(a) the suitability/effectiveness of automated systems to identify content or accounts likely to violate different provisions within your terms of service, including the factors that materially impact suitability/effectiveness (e.g. language of content, type of content) including: **(i) the suitability/effectiveness of automated systems to take down content, apply access restrictions or ban accounts in relation to any or certain provisions within your terms of service without further assistance from human moderation;**
Response:

(ii) how you use your recommender systems to restrict access to certain content, and how you measure the effectiveness and any unintended consequences of using the recommender system in this way;
Response:
(iii) whether and how automated moderation systems differ by type of content (e.g., audio, video, text) or type of violation (of provisions within your terms of service) and any relevant information about costs of these different systems;
Response:
(iv) how data is used to develop, train, test or operate content moderation systems is sourced for different provisions within your terms of service;
Response:
(v) how performance/effectiveness/accuracy of automated systems are assessed and improvements then made, including any relevant considerations or differences for different provisions within the terms of service (e.g., tolerance level for false negatives and false positives between different provisions);
Response:
(vi) how and when automated systems are updated, and the trigger for this (e.g., in response to changing user behaviour or emerging harms);
Response:
(vii) what safeguards are employed to mitigate biases or adverse impacts of automated content moderation (e.g., on privacy and/or freedom of expression), and any relevant considerations or differences for different provisions within the terms of service;
Response:
(b) the range and quality of third-party content moderation system providers available in the UK, particularly for different provisions within your terms of service;
Response:
(c) the process and costs associated with expanding use of existing automated moderation systems for additional provisions in your terms of service, and any relevant barriers or challenges in deploying these automated moderation systems or expanding or upgrading these systems to cover new or additional provisions;
Response:
(d) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 13: How do you use human moderators to enforce terms of service provisions about taking down or restricting access to content, or suspending or banning accounts? **In your response to this question, please provide information relating to (a) – (c).**
Response:
(a) how you determine your services' resource requirements in relation to human moderation, and the factors (or key factors) that impact these requirements (e.g., increases in content or users, the range or types of content prohibited in your terms of service or technological advances in your automated system) including; **(i) which languages are covered by your moderation team and how you decide which languages to cover;**
Response:
(ii) whether moderators are employed by the service or outsourced, or are volunteers/users and any differences regarding how different provisions within the terms of service are moderated;
Response:
(iii) whether and how moderators are vetted, and any relevant consideration for how moderators are assigned to different roles relating to different provisions within the terms of service;
Response:
(iv) the type of coverage (e.g., weekends or overnight, UK time) moderators provide and any relevant considerations for different provisions within the terms of service;
Response:
(b) the process and costs associated with extending the use of human moderation for new/additional provisions in your terms of service, and any relevant barriers or challenges to adding new/additional provisions in your terms of service in relation to your human moderation resources;
Response:
(c) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 14: What training and support is or should be provided to moderators, and what are the costs incurred by providing this training and support? **In your response to this question, please provide information relating to (a) – (g).**
Response:

(a) whether certain moderators are specialised in certain harms or subject material relating to different provisions in the terms of service;
Response:
(b) how services can/should/do assess the accuracy and consistency of human moderation teams;
Response:
(c) the impact of mental health or well-being support for moderators on the effectiveness of content moderation (including impacts on turn-over in moderation teams);
Response:
(d) whether training is provided and/or updated (including for emerging harms), and the frequency of these updates;
Response:
(e) the costs of creating training materials and support systems, and then the costs of updating or expanding these materials and systems (when relevant/required);
Response:
(f) how training, guidance and/or any relevant support systems and/or materials are provided to moderators including which moderators it is provided to (internal, contract, volunteer etc);
Response:
(g) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 15: How do human moderators and automated systems work together, and what is their relative scale in relation to each other regarding how you ensure your terms of service are enforced? **In your response to this question, please provide information relating to (a) – (e).**
Response:
(a) how and when automated systems or human moderators are deployed in the moderation process;
Response:
(b) the costs of different systems or processes and of using different combinations of these systems and processes. In the absence of specific costs, please provide indication of cost drivers (e.g., moderator location) and other relevant figures (e.g., number of moderators employed, how many items the service moderates per day);
Response:

(c) how the outputs of human moderators, or appeal decisions are used to update the automated systems, and what steps are taken to mitigate bias;
Response:
(d) whether there are any relevant differences or considerations for costs or quality assurance processes for moderating different provisions within the terms of service; and
Response:
(e) any other information.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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

For providers of online services

Question 17: What, if any, methods are in place for identifying, defining or categorising content as journalistic content, content of democratic importance or news publisher content on your service? In particular, please provide any evidence regarding the effectiveness of any existing methods.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 18: Moderating journalistic content, news publisher content and content of democratic importance

For providers of online services

Question 18: What considerations are taken into account when moderating journalistic content, news publisher content and content of democratic importance? In your response to this question, please provide information relating to (a) – (e) where relevant.
Response:
(a) once identified, how journalistic content, news publisher content and content of democratic importance is actioned and what kind of action is taken; and how that differs from the moderation of other types of content
Response:
(b) the factors that are or should be considered when taking action (e.g.: downranking/removal/suspension/ban or other) regarding this content
Response:
(c) the proportion of all journalistic content, content of democratic importance and news publisher content actioned upon by you that is actioned based on algorithmic decision making
Response:
(d) the proportion of all journalistic content, content of democratic importance and news publisher content actioned upon by you that is reviewed by human moderators and on what basis content is escalated to be reviewed by human moderators
Response:
(e) any insights into the costs of moderating journalistic content and content of democratic importance, including set up and ongoing costs in terms of employee time and other material costs.
Response:
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:
(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:
(a) any known impacts of over-removal or erroneous removal of news publisher content, journalistic content or content of democratic importance
Response:
(b) briefings, investigations, transparency reports, media investigations and research papers regarding misuse of such speech protective provisions
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

For providers of online services

Question 21: What are the current complaints, counter-notice or other appeal processes for users to contest any action taken by you regarding journalistic content, news publisher content and content of democratic importance on your service? In your response to this question, please provide information relating to (a) and (b) where relevant.
Response:
(a) any initiatives taken to create and increase awareness about the process for users to complain and/or appeal content removals
Response:

(b) any measures currently in place to prevent individual or systematic misuse of any protections for news publisher content, journalistic content or content of democratic importance.
Response:
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 providers of online services

Question 22: Do you carry out any internal impact assessments to understand the freedom of expression and privacy implications of existing policies regarding journalistic content, news publisher content and content of democratic importance? In your response to this question, please provide information relating to (a) and (b) where relevant.
Response:
(a) explain which elements of your service design or operation they relate to and which factors they take into account
Response:
(b) provide relevant briefings, investigations, transparency reports, media investigations and research papers.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 23: What, if any, measures are 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:
(a) whether there are any additional measures/safeguards that are put in place during local or national elections.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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:
(a) whether there are any additional measures/ safeguards that can be put in place during local or national elections
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – User empowerment duties

Question 25: Detecting and moderating relevant content

For providers of online services

Question 25: What processes do you use to detect relevant content and how do you moderate it? In your response to this request, please provide information relating to (a) – (g) where relevant.
Response:
(a) what systems you use for detection
Response:
(b) further to the above, if there are any important features that you take into account to make distinctions between content, e.g. features that might identify a piece of content as promotional suicide material versus content intended to support users at risk of suicide
Response:
(c) where distinctions are made, the extent to which content is actioned automatically, by human moderation, through user reports, other methods or a combination of methods
Response:
(d) any insight into the cost of these processes, including set-up and on-going costs, in terms of employee time and any other material costs
Response:
(e) whether relevant content is allowed or prohibited on your service
Response:
(f) whether you measure the incidence of users encountering such content, and if yes, whether these systems are different to those measuring other types of content, including illegal content

Response:
(g) if you offer users separate complaints procedures for moderated legal content versus illegal content, how often users report content through these channels, and what proportion of content is removed following a complaint
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 26: Impact of relevant content

For all respondents

Question 26: Can you provide any evidence on whether the impact of relevant content differs between adults and children on user-to-user services? We are interested in particular in 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 27 and 28: Experience of specific types of users

For all respondents

Question 27: Can you provide evidence around the types of adult users more likely to encounter relevant content, and the types of adult users more likely to be affected by such content?
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

For all respondents

Question 28: How do you consider the experience of users who have a protected characteristic, or those considered to be vulnerable or likely to be particularly affected by certain types of content? In your response to this request, please provide information relating to (a) – (c) where relevant.
Response:
(a) what criteria you use to determine whether a user is vulnerable or likely to be particularly affected by certain types of content, or if you do not categorise users as vulnerable and why
Response:

(b) if your service collects any information about users that could be used to identify them as having a protected characteristic, vulnerable or likely to be particularly affected by certain types of content and, if so, what information you collect
Response:
(c) if you conduct any research into the experience of the above users on your service
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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:
(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:

For providers of online services

Question 30: How do you design features to enable adult users to have greater control over the content they encounter, when are they offered to users, and what are the broader impacts on your system in deploying them? (For the purposes of our evidence base we are interested in features that enable control over a range of content, not solely **relevant content**). In your response to this request, please provide information relating to (a) – (d xi) where relevant.

Response:
(a) how you measure and what evidence you can provide around the effectiveness of these features in terms of achieving their respective aims to prevent adults from encountering content that they do not want to see
Response:
(b) how you measure user engagement with these features, and any evidence you can provide around this
Response:
(c) how you ensure that these features are suitable for all adult users and that they're easy to access, including considerations for users with protected characteristics and/or vulnerable users
Response:
(d) how you decide when to offer users these features, or how to present the use of these features to users. This includes but is not limited to the following aspects, i) – xi).
Response:
i) how you develop the user need for these features, and the factors considered when determining to develop them
Response:
ii) whether these features are on by default, and in what circumstances
Response:
iii) whether these features are personalised for specific types of users
Response:
iv) when to offer users these features
Response:
v) whether, when or how often to remind users of these features - this can mean reminding users to make an initial choice, or checking if a user wants to update the initial choice later on (and if so, how frequently)
Response:
vi) where users learn about these features
Response:
vii) how to provide information about these features, including the level of detail and the words used to describe complex or technical concepts
Response:
viii) whether users have choice of controls over specific types of content
Response:
ix) how you decide whether to iterate, replace or keep such features
Response:

x) any other factors not already covered above that you take into account when considering such features
Response:
xi) any insight into the cost of these features, including set-up and on-going costs (in terms of employee time and any other material costs) as well as any intended and unintended impacts on the service more broadly (e.g., the technical feasibility of implementing filter tools, or reducing functionality based on verification status).
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – User identity verification duties

Question 31 and 32: Circumstances where user identity verification is offered and how

For all respondents

Question 31: What kind of user-to-user services currently deploy identity verification and in what circumstances? In your response to this request, please provide information relating to (a) – (c) where relevant.
Response:
(a) the ways in which these identity verification methods are beneficial, both to the user and to the service
Response:
(b) what documentation you understand to be necessary for different types, or levels, of identity verification on user-to-user services
Response:
(c) whether you believe there are there any other circumstances where identity verification should be offered on user-to-user services.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

For providers of user-to-user services that provide some types of identity verification for individual adult users

Question 32: In respect of the identity verification method(s) used on your service, please share any information explaining:

(a) in what circumstances identity verification is offered on your service and why, and to which category/categories of users
Response:
(b) what evidence and steps are taken to verify the identity of a user, e.g., which attributes are checked, what aspects of verified users are known only to the provider and what aspects are made available for other users to see, including whether processes regarding adult users are different to those regarding children
Response:
(c) whether the process is, or can be, tailored to users in different geographical areas, such as the UK
Response:
(d) whether you engage third party providers to provide all or part of this identity verification process and, if so, which providers
Response:
e) once a user has their identity verified, what this allows them to do on your service, and if relevant, what activities this enables on another service
Response:
f) how your identity verification policies have been developed, including any research that you can share
Response:
g) any steps you take to ensure that identity verification is available to all adult users, including users who may not be able to access certain types of identity verification
Response:
h) any consideration around users who may be vulnerable participating in the identity verification method
Response:
i) how you manage the identity verification of users who have multiple accounts
Response:
j) how you manage different identity verification methods operating simultaneously on your service, such as forms of age verification that require ID to complete the process, monetised schemes and notable user schemes, and how you consider user perceptions of these different methods
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 33: Cost and effectiveness of these methods

For all respondents

Question 33: Please share any information about the costs and the effectiveness of identity verification methods In your response to this request, please provide information relating to: - (a) – (d) where relevant for all respondents, and - f) and g) where relevant for providers of user-to-user services that provide some types of identity verification for individual adult users.
Response:
(a) any insight into the cost of identity verification methods, including set-up and on-going costs, in terms of employee time and any other material costs, as well as any intended and unintended impacts on services more broadly
Response:
(b) how effective these identity verification methods are in verifying the identity of a user for the particular purpose for which verification is carried out
Response:
(c) any other benefits or unintended consequences from these schemes existing
Response:
(d) the safeguards necessary to ensure users' privacy is protected
Response:
For providers of user-to-user services that provide some types of identity verification for individual adult users
(e) any unintended consequences of implementing identity verification, such as the impact this may have on your site's ecosystem
Response:
(f) how you envisage your service operating in the digital identity market, bearing in mind moves towards cross-industry and federated identity schemes
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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: Children are aware that the content or sources that they engage with online are not always reliable, or do not always reflect additional perspectives. Qualitative research commissioned by Ofcom heard from one child who said when she thought a piece of content that was not genuine, she would verify such content based on the number of followers or likes they had.[45] The child also said she felt content by "more famous people" would be more likely to be true. In this regard, user identity schemes may give children more confidence that the content they are seeing is reliable and well founded. However, user identity verification can also be used by malign actors to impersonate or mislead users. Reporting on X (formerly Twitter) illustrated that its verification scheme is vulnerable to scams[46] and has led to an explosion of verified bot accounts.[47] This could leave children susceptible to commercial risk – such as fraud. Further, research by Ofcom[48] also showed that nearly a quarter (23%) of children claimed to be confident in their ability to recognise whether a profile was genuine could not recognise a fake social media profile. One child had said the verified 'tick' was a key indicator for them whether content was genuine – which in the case of X could put them at significant risk. If service providers choose to incorporate identity verification for children into their service, they should consider that it is rare for children to have officially-recognised identity documents (e.g. drivers licenses, online banking). Any processing of children's data must adhere the 15 standards of the ICO's Age Appropriate Design Code,[49] including collecting the minimal amount of data to confirm the child is who they say they are and ensuring this data is not shared on or used for anything else outside of its original purpose. The process for identity verification should be clearly explained to users and information on how users can register, as well as information on how their data will be used, should be clearly reflected in published terms that meet standards in the IEEE Std. 2089[50] to present them in an age-appropriate way that is accessible to all children – **see Q1**.
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No

For providers of user-to-user services that provide some types of identity verification for individual adult users

Question 35: How do you measure engagement with your identity verification methods? In your response to this request, please provide information relating to (a) and (b) where relevant.
Response:

[45] Revealing Reality & Ofcom (2024) *Children's Media Lives 2024: Ten years of longitudinal research*, p. 40
[46] Burgess, M. (2022) Elon Musk's Twitter Is a Scammer's Paradise. *The Wired*
[47] Perez, S. (2024) It sure looks like X (Twitter) has a Verified bot problem. *TechCrunch*
[48] Ofcom (2023) Children and Parents: Media Use and Attitudes, pp. 42-43
[49] ICO, Age appropriate design: a code of practice for online services
[50] IEEE SA, Std. 2089

(a) take-up of identity verification by your users
Response:
(b) any insight into whether identity verification has any other effect on user behaviour, such as the content that users post and the amount that they engage with your service.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – Fraudulent advertising

Questions 36 – 42: Overarching considerations

For all respondents

Question 36: Please provide evidence of the following:
(a) The most prevalent kinds of fraudulent advertising activity on user-to-user and search services (e.g. illegal financial promotions, misleading statements, malvertising)
Response:
(b) The harms associated with different kinds of fraudulent advertisements, the severity of such harms, and, if relevant, how this varies by user group
Response:
(c) The key challenges to successfully detecting different types of fraudulent paid-for advertising, and how these challenges can be minimised or resolved
Response:
(d) The prioritisation of suspected fraudulent advertising within all categories of harmful advertising queues, e.g. account verification, user reports, appeals
Response:
(e) The proportion of fraudulent advertisements that are currently estimated to remain undetected by services' systems.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 37: What technological developments aiding the prevention/detection of fraudulent advertisements do you anticipate in the coming years, and how costly and effective do you expect them to be? What are the challenges/barriers to their development?
Response:

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

Question 38: If you have information/evidence/suggested mitigations to share which may be useful in the preparation of codes of practice, which is not covered by the questions above, please include these under 'Overarching considerations'.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

For providers of online services

Question 39: What proportion of all paid-for advertising on your service is identified as fraudulent advertising?
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 40: Does your service take any steps to warn users of the risk of encountering fraudulent advertising or to educate them about how to identify potentially fraudulent advertising?
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 41: Please provide information regarding the proportion of successfully identified fraudulent advertisements that are identified via:
(a) automated systems
Response:
(b) human processes
Response:
(c) user reports
Response:
(d) other (please provide further detail).
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 42: What is the average and/or median time taken between the identification of a fraudulent advertisement and its removal/other actions taken? (If other actions taken, please specify what they are).
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 43: Proactive technology

For all respondents

Question 43: Please provide any evidence you have regarding proactive technologies which could be used to identify fraudulent advertising activity. In particular, we are interested in information related to the following points:
(a) The kinds of proactive technology which are/could be applied to identify or prevent fraudulent advertising
Response:
(b) A brief description of how these technologies are/could be integrated into the service
Response:
(c) The effectiveness, accuracy and lack of bias of such technology (including compared to alternative proactive and non-proactive methods) in relation to detecting fraudulent advertising and accounts which post fraudulent advertising material
Response:
(d) How proactive technologies are maintained and kept up to date
Response:
e) Information related to the associated time and/or costs for set-up, operation, and human review
Response:
f) The cost of integrating such technologies: (a) for the first time; and (b) when updating these technologies over time
Response:
g) Whether there are cost savings associated with these technologies
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 44: Advertising onboarding and verification

For all respondents

Question 44: Please provide any evidence you have regarding the processes for advertiser onboarding and verification related to protections against fraudulent advertising. In your response, please indicate whether these processes are currently implemented in respect of services which are in scope of the Act or whether they stem from another sector In particular, we are interested in information related to the following points:
(a) The criteria which advertisers are verified against, including documentation/evidence used to support verification, and what advertisers are required to declare
Response:
(b) The role of (a) automated processing and (b) human processing in the verification process, and how they interact
Response:
(c) The costs associated with advertiser verification and how those costs vary as scale increases
Response:
(d) The percentage of advertiser accounts that are verified
Response:
e) Whether advertisers are permitted to publish advertisements on the service while the verification process is ongoing
Response:
f) Whether there are additional/specific verification checks for advertisers placing adverts of certain kinds or targeting certain audiences, such as about specific products or services, or targeting users under the age of 18
Response:
g) Whether the verification of an advertiser account expires after a certain amount of time or certain activity, such as when advertisers make changes to their account or profile
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 45: Service review of submitted advertisements/sponsored search results

For all respondents

Question 45: Please provide any evidence you have regarding the processes that services in scope of the Act have in place to review submitted paid-for advertisements and identify fraudulent advertising material. In particular, we are interested in information related to the following points:

(a) The percentage of submitted advertisements which are reviewed both (i) prior to and (ii) after publication
Response:
(b) The role (i) automated processing and (ii) human processing play in the review process and how they interact
Response:
(c) The red flags which trigger advertisement review processes both (i) prior to and (ii) after publication and the basis on which those red flags are selected
Response:
(d) The timescales for review
Response:
(e) What happens to the advertisement's visibility and reach, if it is flagged as suspected as being fraudulent (either by a user or automated system)
Response:
(f) The costs associated with the review of submitted paid-for advertisements
Response:
(g) Whether trusted flagger reporting is employed to inform services' review processes. If it is, how is it applied, what guidelines / criteria does it follow, and who are those trusted flaggers?
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 46: Advertiser appeals of verification/review decisions

For all respondents

Question 46: Please provide any evidence you have regarding advertiser appeals of verification/review decisions relating to fraudulent advertising on services in scope of the Act. In particular, we are interested in information related to the following points:
(a) The role of (i) automated processing and (ii) human processing in the appeals process, and how they interact;
Response:
(b) The level of proof required for an appeal to be accepted;
Response:
(c) The most frequent bases for appeals against sanctions decisions on fraudulent advertising content
Response:

(d) The ratio of decisions that are appealed against
Response:
(e) The costs associated with appeals
Response:
(f) The proportion of appealed decisions which are upheld and overturned
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 47: User reporting mechanisms

For all respondents

Question 47: Please provide any evidence you have regarding user reporting mechanisms for fraudulent advertising on services in scope of the Act. In particular, we are interested in information related to the following points:
(a) What user reporting tools there are for paid-for advertisements, and how these tools differ from those for user-generated content and/or search results and other search functionalities that are not paid-for advertising
Response:
(b) What percentage of user reports of advertisements relate to suspected fraudulent content, and the processes for taking action in relation to such reports
Response:
(c) Any statistics you can share on (i) the number of user reports of suspected fraudulent advertising received and resolved over a specific period and (b) the number of initial decisions appealed by users who made the report
Response:
(d) The criteria used to classify and prioritise user reports
Response:
(e) The median and/or average time it takes to respond to a user report, and any measures that are in place to ensure timely and accurate responses to user reports
Response:
(f) Any measures taken to make user reporting tools accessible, easy to use and easy to find for users
Response:
(g) How transparency and communication is maintained with users who have submitted reports

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

Question 48: Use/involvement of third parties

For all respondents

Question 48: Please provide any evidence relevant to fraudulent advertising that you have, regarding the involvement and role of third parties in the provision of paid-for advertisements on services in scope of the Act. In line with the proportionality criteria under sections 38(5) and 39(5) of the Act, we welcome information related to how the involvement of third parties impacts the degree of control that services have over fraudulent advertising content. We also welcome information regarding contractual arrangements and how those arrangements are enforced.
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 49: Generative AI and deepfakes

For all respondents

Question 49: Please provide any evidence you have regarding the impact of generative AI developments and deepfakes on the incidence and detection of fraudulent advertisements on services in scope of the Act. In particular, we are interested in information related to the following points:
(a) The frequency of deepfake fraudulent advertisements' occurrence, in absolute terms and/or as a proportion of all fraudulent advertisements, and how you expect this to evolve in the future
Response:
(b) What methodologies/technologies are currently employed to detect fraudulent advertisements which include deepfake or otherwise AI-generated content, and the effectiveness of these tools
Response:
(c) Whether detection technologies are developed in-house or acquired from a third-party, and how long it takes to develop and/or integrate those tools into wider systems
Response:
(d) The accuracy of detection methods, including true positive and false positive rates

Response:
(e) The costs associated with the development/acquisition and deployment of these detection mechanisms
Response:
(f) The types of deepfake or AI-generated content (in terms of either media type or subject) in fraudulent advertisements that are most difficult to detect i) via automated processes, ii) by human moderators, iii) by service users
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Your response – Access to information about a deceased child's use of a service

Questions 50 – 55: Processes for requesting information about a deceased child's use of a service

For all respondents

Question 50: What kinds of information might parents want to see about their child's use of the service?

Response:

To support Ofcom's understanding of how this measure should work, we have consulted with the Bereaved Families for Online Safety (BFOS) who campaigned strongly for these measures. They told us they believed the following information may be of use to other parents in these situations:

- **If a child had access to the service (e.g. was age verification in place or a bar which would have meant they could not have accessed it)**.
- **If a child used that service and** to what extent – including what username they had, how many accounts they had.
- **If a child visited any advertisements for products on the service** – and if so, what products they were.
- **If a child had been taken from one service to another** – including if certain content or activities had transferred inter-service.
- **Information about what/who a child interacted with**, with personal data redacted before sharing. Should this reveal anything, the tech company/police could investigate further.

With respects of this power, we recommend that Ofcom produces guidance for tech companies which helps to balance children's safeguarding with their rights, and/or have a dedicated individual knowledgeable of children's rights to provide a trusted and accountable procedure by which to

assess and determine children's best interests in relation to the digital environment.[51] In addition, companies should:

- **Take appropriate action to respect and protect children's rights** by developing proposals to mitigate potential infringement of their rights.
- **Consult with children and experts** when operating due diligence processes, using evidence-based research recording the decision-making process – particularly where rights appear to be in conflict.
- **Outline and review rights conflicts**, such as when children's rights come into contention with corporate or business interests.

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

Response: No

Question 51: How long should it take to receive information in response to a request?

Response:

When making a request, it is critical that communication is open and transparent from the tech company with parents. An arbitrary timeframe may not be suitable depending on the complexity and amount of data/activity that a child has on the platform. If there is a lot of sensitive data, it may also take time to redact.

The Bereaved Families for Online Safety (BFOS) also told that it is crucial that responses for requests are received in a timely manner and that such timeframes are indicated and communicated clearly. Delivery of this information should not be from a bot and should be from a human actor. Parents also understand that, depending on the type of information requested, a single, specific length of time may not be ideal. For example, a request about whether or not a child had access may require less time than accessing information relating to children's metadata.

The BFOS recommended:

- **An ombudsman-style approach** which clearly outlines specific timeframes, and ensuring these are written down. As an example, the regulator Ofwat indicates on their website the timeline for investigations depending on the type of complaint or request.[52] Bereaved families also want to see tech companies adhere to similar timelines as other regulated companies in different sectors.
- **A response in the first couple of days from a human actor**, recognising the fact that bereaved families have lost a child.

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

Response: No

[51] Livingstone, S., Cantwell N., Özkul, D., Shekhawat, G. & Kidron, B. (2024) *The best interests of the child in the digital environment*. Digital Futures for Children, LSE & 5Rights Foundation, pp.
[52] Ofwat (2016) Our timeframes for handling cases

Question 52: What mechanisms could, or should services provide for parents to find out what they need to do to obtain information and updates in these circumstances?

Response:

The Bereaved Families for Online Safety (BFOS) told us that having to navigate complex systems adds to the pain of processing the loss of a child in the most catastrophic of circumstances, and that the process is often intimidating. Indeed, finding similar mechanisms – i.e. reporting – is difficult to do, and not on the frontpage of a service, which restricts the abilities of parents to raise their concerns. When speaking with bereaved families, one parent told us that, in order to report a harmful piece of content on TikTok, she had to Google how to do it. It was only through working with a journalist that it was taken down from the service.[53]

When making a request for children's data, mechanisms must be available that clearly explain and assist parents in the process, as well as what tech companies are required to do, to ensure the process is transparent and as smooth as possible. Without appropriate mechanisms in place, it will be impossible for parents to build up trust and confidence in using systems that allow them to request.

The BFOS recommended that:

- **Services should have a dedicated contact page which provides a number to someone who can support them.**
- **A dedicated open forum or chat room for parents to go to**, allowing parents to ask questions and communicate with services.
- **A page on Ofcom's website that outlines what to expect from this process**, as well as what the expectations are of the service when communicating with parents individually.
- **A caseworker at Ofcom** that acts similarly to a family liaison officer to explain, advise, and guide families through the process, giving them the ability to feedback. Additionally, being issued a specific case number when making a request, complaint, or appeal may make it easier for families to follow-up on their cases.

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

Response: No

Question 53: What support or information do parents need to guide them through the process of making a request?

Response:

The Bereaved Families for Online Safety (BFOS) spoke to us about how difficult they found it to speak to humans at tech companies, particularly early in the process of reporting. This made relatively obscure systems even more difficult to navigate. Automated contact routes were found by parents to be longwinded and unresponsive. Bereaved families want the ability to be able to reach a human point of contact to explain their circumstances, without having to constantly re-tell lots of people.

[53] Blackburn, R. (2024) TikTok challenges: tech site 'removes' dangerous breathing dares after NationalWorld campaign. *NationalWorld*

Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No

For providers of online services

Question 54: What kinds of information do you provide and how do you provide this information? In your response to this request, please provide information relating to (a) where relevant.
Response:
a) **If there are certain types of information you cannot provide, please explain why, for example whether there are technological, cost or privacy factors that mean certain kinds of information may not be feasible to provide**
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Question 55: How long does it typically take you to provide information in response to a request? In your response to this request, please provide information relating to (a) where relevant.
Response:
a) How long should it reasonably take services to provide information in these circumstances?
Response:
Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Questions 56 and 57: Complaints systems

For all respondents

Question 56: What can providers of online services do to ensure the transparency, accessibility, ease of use and users' awareness of complaints mechanisms in relation to deceased user information request processes?
Response: It is crucial that, in order for parents to feel confident using complaints systems, that tech companies operate them in a way that is easy to use, understandable, and is clear about how the process and timeline for submitting a complaint works. The Bereaved Families for Online Safety (BFOS) told us repeatedly that conventional reporting and complaints systems are too onerous, and that in some cases they had felt "gaslit" by these

processes. Complaints mechanisms in relation to obtaining children's data must be usable and sensible, and always actioned where a request is appropriate.

Parents also told us that they valued the ability to redress and challenge systems – which they frequently find is often difficult to do in other mechanisms – e.g. reporting. Where a parent seeks to complain regarding a request for their child's access to data, a timeframe for handling the complaint must be clearly communicated, and any updates to their complaint must be delivered by a human and not an automated system. Parents must also be given a clear route to redress if their request was turned down.

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

Response: No

For providers of online services

Question 57: Can you provide any evidence or information about the best practices for effective complaints mechanisms which could inform an approach to complaints about information request processes pertaining to a deceased user?

Response:

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

Response:

Question 58: Evidence

For providers of online services

Question 58: What kinds of evidence do you require about the identity of the person making the request and their relationship to the deceased user?

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

Response:

(a) Do you, or would you, require different kinds of evidence in the event that the deceased user is a child?

Response:

(b) What evidence do, or would, you require that a user is deceased?

Response:

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

Response: