Third phase of online safety regulation

Additional duties for categorised services

Welsh version available

Call for evidence

Published 25 March 2024
Closing date for responses: 20 May 2024

Contents

Section

Annex

1. Overview

1.1 Ofcom is the UK's communications regulator, overseeing sectors including telecommunications, post, broadcast TV and radio, and online services. We were appointed the online safety regulator under the Online Safety Act 2023 (the Act) in October 2023.

1.2 The Act makes companies that operate a wide range of online services legally responsible for keeping people, especially children, safer online. It introduces a system for categorising some regulated online services based on key characteristics, including user numbers and functionality. Some regulated services will be designated as category 1, 2A or 2B services if they meet certain thresholds set out in secondary legislation.[1] Once the secondary legislation is passed, Ofcom is required to develop and produce a published register of categorised services (and list of emerging category 1 services).[2] If the secondary legislation is consistent with our advice to the Secretary of State, also published today, based on preliminary analysis we expect to categorise between 35-60 services across the three categories - although the numbers may change following the designation process. All other services – i.e., the vast majority of services in scope of the Act - will not be categorised and will not be subject to any of the additional duties described below.

1.3 Providers of these categorised services will be required to comply with some additional duties – depending on the category they fall within. Figure 1.1. below summarises the duties that apply to the different categories.

Figure 1.1: Additional duties that apply to categorised services

Categories →	Category 1	Category 2A	Category 2B
Categorised services must comply with additional duties relating to the below:			
Transparency reporting	✅	✅	✅
Enhanced requirements on risk assessments and record keeping	✅	✅	
Additional terms of service duties	✅		
Protections for news publisher and journalistic content, and content of democratic importance	✅		
Providing user empowerment features	✅		
Providing user identity verification options	✅		
Prevention of fraudulent advertising	✅	✅	
Disclosure of information about use of the service by a deceased child user	✅	✅	✅

[1] The Act required Ofcom to produce advice for the Secretary of State on where the thresholds should be set, based on research we have carried out.

[2] Based on current estimates, we expect to produce a published register of categorised services (and list of emerging category 1 services) by the end of 2024. Timings may however be subject to change as the passing of secondary legislation may be impacted by a General Election.

1.4 The additional duties for categorised services play an important role in advancing the strategic goals of the online safety regulatory regime. They are intended to lead to greater choice and transparency for users including children, giving them more meaningful control over their online experiences. There are specific provisions to tackle online fraud and empower users to protect themselves from it. Core protections will be afforded to news publisher content, journalistic content and content of democratic importance which are fundamental to democratic discourse.

1.5 Ofcom is taking a phased approach to implementing the Act. The first phase relates to illegal harms and we recently consulted on the codes of practice and guidance for these duties. We currently plan to publish our final codes and guidance before the end of 2024. Phase two relates to the protection of children and we plan to consult on these codes of practice and guidance in Spring 2024. **This call for evidence continues Ofcom's work to implement the online safety regime, and focuses on the additional duties providers of categorised services will need to comply with, which forms part of phase three of implementation.**

1.6 As part of phase three, we will follow a three-step process. First, service providers subject to additional duties will be identified, which is a process we describe in more detail below in section 2. Second, we will consult on draft codes and guidance which will detail how services can comply with these additional duties. We will do this in two parts, first consulting in Summer 2024 on our guidance relating to our transparency reporting regime to prioritise its swift implementation in 2025. This will be followed by a further consultation in early 2025 on further additional duties for categorised services. This document calls for evidence to support the early 2025 consultation. Third, following these consultations, we will publish final codes and guidance.[3]

Summary of duties in the Act and the evidence we are seeking

Under the Act, Ofcom must produce a range of codes of practice and guidance outlining steps that companies may take to comply with their new duties.

We are seeking evidence from stakeholders to inform the codes of practice and guidance that Ofcom must produce to implement the third phase of online safety regulation.

Additional terms of service duties: duties on category 1 service providers to use proportionate systems and processes to ensure that taking down or restricting access to content, and suspending or banning users, is only carried out in accordance with the terms of service related to these actions. Providers must also ensure the terms of service related to these actions are consistently enforced. We want to know how:

- providers of online services ensure their terms of service enable users to understand when and how different types of enforcement action will be taken against content or accounts; they measure whether users read, understand, and follow their rules; and they avoid over- or underenforcing their terms of service when implementing moderation systems.

[3] Timings may be subject to change as implementation of some parts of the regime is dependent on the passing of secondary legislation, which may be impacted by a General Election which will be called at some point before January 2025. We have not factored a General Election into our planning to date.

Protections for news publisher content, journalistic content and content of democratic importance: duties on category 1 service providers to use processes aimed at protecting these types of content. We want to know about:

- the identification, classification, and moderation of the above content on in-scope services; the effectiveness and cost implications of these processes; and measures used to prevent the misuse of systems to identify and categorise this type of content including the current application of complaints and appeals processes.

Providing user empowerment features: duties on category 1 service providers to give adult users the option to reduce the likelihood of being exposed to certain types of content and to filter out non-verified users.[4] We want to know about:

- the detection, classification and moderation of relevant content on in-scope services; the tools and features that are currently offered to empower users to navigate relevant content and interactions between different types of users, including the effectiveness, take-up and costs associated with these tools and features; the incidence of relevant content for adult users across in-scope service; and the experience of adult users with a protected characteristic in encountering relevant content, or of those likely to be particularly affected by such content.

Providing user identity verification options: duties on category 1 service providers to offer UK adult users the option to verify their identity. We want to know about:

- the circumstances where identity verification is offered on user-to-user services, how this is done, and the cost and effectiveness of these methods; and the broader implications surrounding identity verification, including user attitudes towards verified accounts and user attitudes to widespread implementation.

Prevention of fraudulent advertising: duties on category 1 and 2A service providers to have proportionate systems and processes in place to protect users from certain types of fraudulent advertising on their service.[5] We want to know about:

- the processes and mechanisms (including their effectiveness, the costs, and the risks of unintended effects) that in-scope services currently use to support both the delivery of advertising and detection of fraudulent advertising material; suggested additional processes and mechanisms that could be implemented in order for relevant services to meet their duties in relation to fraudulent advertising; and any relevant evidence regarding the role of third-party intermediaries involved in the process of serving ads on in-scope services and their relationship to those services.

Access to information about a deceased child's use of a service: duties on category 1, 2A and 2B service providers to include in the terms of service their policy on requests from parents, or those with parental responsibility for a child, for information about their deceased child's use of the service and provide dedicated mechanisms for parents to engage with the service provider in those circumstances. We want to know:

[4] Providers of online services likely to be accessed by children must also ensure children enjoy greater protections from pornographic content and other types of content that is harmful to them, such as promotional suicide or eating disorder content. We published our call for evidence on the protection of children provisions as part of our second phase of online safety regulation on 10 January 2023 and will be consulting on our codes of practice and guidance relating to protection for children in Spring 2024.

[5] The fraudulent advertising provisions are the only advertising-specific duties in the Act.

- what kinds of evidence services might require about the parent's identity or relationship to the child, and about the death of the child; what kinds of information parents might request about their child's use of the service, what information services do or might provide and how, and the challenges or trade-offs of doing so; how long it should reasonably take services to provide that kind of information; and what mechanisms currently exist for parents to find out what they need to do to obtain information and updates in these circumstances, whether these are easy to use, and what other mechanisms might be made available.

We are seeking evidence and input from any interested stakeholders on the questions set out in this document to inform the development of regulation from the outset. We also welcome any additional information, beyond that identified in the questions, that stakeholders may consider is relevant to Ofcom preparing codes of practice and guidance on these policy areas. This may include relevant information on how stakeholders are preparing for regulation in other (non-UK) jurisdictions. We recognise stakeholders may have already submitted relevant information to Ofcom on these topics. Respondents may resubmit information to us if they wish, or alternatively simply bring a previous submission to our attention.

We welcome responses from interested stakeholders to the questions set out in sections 3 – 8 in this document.

This call for evidence will close on 20 May 2024.

2. Background

2.1 In this section, we provide a summary of how this call for evidence relates to other parts of the online safety regime. We then summarise the relevant additional duties for categorised services.

The Online Safety Act 2023

2.2 The Act received Royal Assent on 26 October 2023, and makes companies that operate a wide range of online services legally responsible for keeping people, especially children, safer online.

2.3 Providers of online services in scope of the Act, such as social media, messaging services and search engines, must, for example, carry out a suitable and sufficient illegal content risk assessment; make it clear in their terms of service how they will protect users; make it easy to report illegal content and complain, and have particular regard to the importance of protecting users' rights to freedom of expression and privacy when implementing safety measures.

2.4 Providers of online services likely to be accessed by children must also ensure greater protections for children from pornographic content and other types of content that is harmful to them, such as suicide or eating disorder content.

Ofcom's phased approach to implementation

2.5 Ofcom is taking a three-phased approach to implementing the UK online safety regulatory regime:

a) phase one: duties regarding illegal harms;
b) phase two: duties regarding protecting children, including from access to pornography;
c) phase three: additional duties for categorised services.

2.6 We published our approach to implementing the Act in October 2023. This approach summarises the key actions companies must take under the Act and sets out our plans for putting online safety laws into practice, including how we will implement phases one, two and three, as well as our supervision programme and work relating to improving media literacy among UK adults and children.[6] It also outlines the outcomes we expect these new online safety rules to deliver.

2.7 We published our consultation on the illegal harms provisions as part of our first phase of online safety regulation on 9 November 2023, and currently plan to publish our statement before the end of 2024. We published our consultation on guidance for service providers publishing pornographic content on 5 December 2023. We published our call for evidence on the protection of children provisions as part of our second phase of online safety regulation on 10 January 2023. We will be consulting on our codes of practice and guidance

[6] We will publish our consultation on a three year strategy for media literacy in Spring 2024. We expect to publish our statement in Autumn 2024.

relating to the protection of children in Spring 2024. Throughout 2024, we will be working to consider responses to our consultations with the aim of having a number of codes of practice and guidance in place in 2025 to inform online service providers' compliance with the Act.

Phase three: additional duties for categorised services

Implementation of phase three

2.8 Due to the wide range of duties relevant to this phase, as well as the need to consider stakeholder evidence gathered from our phase one and two consultations, we have staggered our calls for evidence and consultations for phase three. We included the transparency reporting regime for in our call for evidence on illegal harms on 6 July 2022 and we published a call for evidence on the research and advice for categorisation on 11 July 2023. We are now publishing this call for evidence on the **additional duties that apply to providers of categorised services**, which we describe below.[7]

The register of categorised services

2.9 The Act introduces a system for categorising online services based on key characteristics, including user numbers and functionality. The providers of categorised services will be required to comply with additional duties depending on which category they fall within (as set out in more detail above).

2.10 Ofcom was required under the Act to carry out research and produce advice to the Secretary of State on the threshold conditions for each category of service. We submitted our advice to the Secretary of State on 29 February 2024 and published our advice on 25 March 2024, alongside this call for evidence. Our advice on how the various thresholds should be set is as follows:

a) **Category 1**: Our advice is that category 1 thresholds should target services that fulfil either of the two following sets of conditions:

i) Condition 1: use a content recommender system; and have more than 34 million UK users on the user-to-user part of the service, representing c.50% of the UK population;
ii) Condition 2: allow users to forward or reshare user-generated content; and use a content recommender system; and have more than 7 million UK users on the user-to-user part of the service, representing c.10% of the UK population.

b) **Category 2A**: Our advice is that category 2A thresholds should target services that fulfil both of the following criteria:

i) Is a search service but not a vertical search service; and
ii) Have more than 7 million UK users on the search engine part of the service, representing c.10% of the UK population.

c) **Category 2B**: Our advice is that category 2B thresholds should target services that fulfil both of the following:

i) Allow users to send direct messages; and

[7] We will consult on online safety transparency guidance in Summer 2024. See paragraph 2.24 and 2.25 below.

ii) Have more than 3 million UK users on the user-to-user part of the service, representing c.5% of the UK population.

2.11 The Secretary of State must consider this advice as part of determining the category 1, 2A and 2B threshold conditions to be set in secondary legislation. Once the secondary legislation has passed, Ofcom will then gather information as needed, including under our statutory powers to request information from regulated services. We will analyse the information gathered about services against the final thresholds and in accordance with the Act produce a published register of categorised services, and a published list of emerging category 1 services.

2.12 The published register of categorised services will determine which companies need to comply with the additional duties in the Act. Assuming secondary legislation on categorisation is finalised by Summer 2024, we expect to publish the register of categorised services by the end of 2024.

Additional duties for categorised services

2.13 The Act makes all regulated online service providers legally responsible for keeping people safer online. The providers of all online services in scope of the Act must put in place measures to protect users from illegal content[8] and to protect children from content that is harmful to them.[9] In addition, providers of categorised services must comply with additional duties.[10] These additional duties are set out below.

Duties for category 1 services only

2.14 There are a range of duties that will only apply to providers of category 1 services. These duties include:

2.15 **Additional terms of service duties:** service providers must use proportionate systems and processes to (a) ensure they act in accordance with their terms of service when taking down or restricting access to content, or banning or suspending users, and (b) enforce those terms of service. Such terms must be clear and accessible, written in sufficient detail, and applied consistently. Users must be provided with reporting and complaints mechanisms in relation to these duties.

2.16 **Protecting news publisher content, journalistic content and content of democratic importance duties**:[11]

[8] For illegal content duties see for example, sections 9-10 (duties on user-to-user services) and sections 26-27 (duties on search services) of the Act. We consulted on the illegal content duties in our illegal harms consultation

[9] For child safety duties see, for example, sections 11-13 (duties on user to user services) and sections 28-30 (duties on search services) of the Act.

[10] There are additional duties on providers of categorised services in relation to record-keeping for the illegal harms risk assessments and the protection of children risk assessments (sections 23(1), 34(9) of the Act – see our consultation on illegal harms.

[11] The Act also places duties on all providers relating to freedom of expression and privacy. These duties require all in-scope services to have particular regard to the importance of protecting users' rights to freedom of expression and privacy when deciding on, and implementing, safety measures and policies (Sections 22 and 33 of the Act).

a) **protecting news publisher content**: service providers must take certain steps before taking action in relation to a recognised news publisher, or in relation to news publisher content.
b) **protecting journalistic content**: service providers must use proportionate systems and processes designed to ensure that the importance of the free expression of journalistic content is taken into account when moderating this content, particularly when considering take-down or user access restrictions and user sanctions for sharing such content, and have a dedicated and expedited complaints procedure in place.
c) **protecting content of democratic importance**: service providers must use proportionate systems and processes designed to ensure that the importance of the free expression of content of democratic importance is taken into account when moderating this content, particularly when considering take-down or user access restrictions and user sanctions for sharing such content, and these apply in the same way to a wide diversity of political opinion.
d) **impact assessments**: service providers must carry out impact assessments relating to users' rights to freedom of expression and privacy, news publisher content and journalistic content.

2.17 **Providing user empowerment features:** service providers must offer adult users features that can be applied to reduce their exposure to certain types of legal content and filter out content from non-verified users. These tools must be easy to access. Service providers will also have to undertake a content assessment to, among other things, measure the incidence of this content on their service.

2.18 **Providing user identity verification**: service providers must offer all UK adult users the option to verify their identity.

Duties for category 1 and 2A services only

2.19 **Prevention of fraudulent advertising:**[12] service providers must have proportionate systems and processes in place designed to prevent users from encountering fraudulent advertisements on the service, minimise the length of time fraudulent advertisements are present on the service, and swiftly remove (in the case of category 1) or ensure that individuals are no longer able to encounter such content in or via search results of the service (in the case of category 2A) when it is alerted to, or becomes aware of, the incidence of such content.

Duties for category 1, 2A and 2B services

2.20 **Access to information about a deceased child's use of a service**:[13] service providers must make clear in their terms of service (or for category 2A services in a publicly available statement) what their policy is about disclosing information to the parents of a deceased child user about the child's use of the service. Service providers must have a mechanism for

[12] User-to-user and search content that amount to a relevant fraud offence is in scope of the general safety duties in the Act; we have consulted on proposed codes of practice.
[13] Separately, Ofcom will have a discretionary power under section 101 of the Act to require service providers to provide information about a child's use of a service, where it is requested by a coroner in connection with an investigation into the death of a child. We will set out our policy for using our section 101 powers in due course.

parents to easily find out what they need to do to obtain information and updates in those circumstances.

2.21 **Transparency reporting:** service providers must publish a transparency report based on notices that Ofcom will issue to providers once a year. We set out these duties and our approach to establishing the transparency regime, including our plan for consultation, below at paragraphs 2.24-2.25. These duties do not form part of this call for evidence.

2.22 The Act does not envisage that all providers of categorised services will adopt the same approach to complying with each of these duties - as in other areas of this regime, it is for providers to determine the most appropriate way to comply with the requirements given the risks they face.

2.23 Ofcom is required by the Act to produce codes of practice and guidance to support industry compliance with these duties. Services will be deemed to comply with the duties if they follow the measures set out in the codes. Guidance is also intended to be an aide to compliance and services would be expected to take guidance into account.

Transparency reporting

2.24 The Act requires Ofcom to establish a service transparency reporting regime. Providers of categorised services will be required to publish a transparency report based on transparency notices that Ofcom will issue to service providers once a year. Ofcom will then publish its own transparency report summarising industry trends and setting out good practice based on service transparency reports and any additional information, such as new research, to help contextualise those findings for the public.

2.25 Targeted transparency reporting requirements are a key tool for driving effective and meaningful change under the online safety regime.[14] In recognition of this, we began engaging with stakeholders early, asking what information would be most useful to include in services' transparency reports and Ofcom's own transparency reports in our illegal harms call for evidence on 6 July 2023. We will consult on our online safety transparency guidance in Summer 2024.

The online safety outcomes we are seeking

2.26 Ofcom's overarching mission is to make communications work for everyone. In the context of our new online safety responsibilities, we aim to ensure a safer life online for people in the UK by ensuring that services take steps to reduce harms and make consumers safer.[15]

2.27 While the onus will be on companies to decide what safety measures they need to apply given the risks their services pose to users, we expect the implementation of the Act to deliver four key outcomes to ensure people in the UK are safer online:

a) stronger safety **governance** in online services,
b) online services are **designed and operated** with safety in mind,
c) greater **choice** for users so they can have more meaningful control over their online experiences, and

[14] Ofcom (2023) Transparency Reporting: The UK Regulatory Perspective.
[15] For more information, see our proposed plan of work for 2024/25

d) users **trust** that they are protected online, including by promoting transparency about services' safety measures and the action Ofcom is taking to improve them.[16]

2.28 The three phases of implementation are fundamental building blocks that together contribute to the delivery of these four outcomes. With the timing of implementation driven by legislation, we have adopted a phased approach. Phases one and two focus on tackling illegal harms and protecting children online respectively and further progress our strategic aims around robust **governance** and **risk assessment** and safer **design** and **operations** within services. This third phase of implementation expands upon that groundwork, with a particular emphasis on enabling **choice**, building users' **trust**, and affording enhanced protections for certain types of content.

2.29 The UK Government's ambition for phase three was to provide UK internet users with additional protections in relation to categorised services, to ensure they apply their policies consistently (only removing or blocking content that is prohibited in clear and accessible terms of service), and are more transparent about their safety measures. Users will be provided with greater choice about the content they see, and with greater control over the people they interact with. Users will also be better informed about the implications of the choices they make. There are specific provisions to tackle online fraud. Core protections will be afforded to news publisher content, journalistic content and content of democratic importance which are fundamental to democratic discourse.

2.30 As set out below, this call for evidence represents an initial step in understanding the changes needed from services to enable these outcomes. We need to understand what can be achieved in this area to recommend proposals on codes and guidance that can deliver the transformative changes we are seeking for users' safety. That is why stakeholder input is so critical at this stage in the process.

Responding to this call for evidence

2.31 We are required to consult widely on any proposals or decisions which impose or amend regulatory obligations. We recognise the importance of close engagement with stakeholders from the very beginning of developing regulation. Therefore, we have been calling for evidence at each stage of our development of the online safety regime.

2.32 We are seeking evidence and input from any interested stakeholders on the questions set out in this document. We also welcome any additional information, beyond that which is specified in the questions, that stakeholders may consider is relevant to Ofcom preparing codes of practice and guidance on these policy areas. This may include relevant information on how stakeholders are preparing for regulation in other (non-UK) jurisdictions.

2.33 Where questions in this call for evidence appear similar to those asked in previous consultations, we highlight the unique policy objectives of the duties on providers of categorised services which are the subject of this call for evidence. Providers of categorised services will represent a subset of the overall regulated industry.

2.34 If stakeholders consider that they have previously given relevant evidence to us on any area covered by this call for evidence, they may resubmit information to us, or alternatively bring a previous submission to our attention with any relevant updates or additions.

[16] See Figure 1, Roadmap in Ofcom's approach to implementing the Online Safety Act - Ofcom.

2.35 If stakeholders think certain evidence is applicable to multiple questions within this call for evidence, please highlight this in your response. We will consider all responses carefully as we start developing our codes of practice and guidance for the third phase of implementing the online safety regime.

3. Additional terms of service duties

The **additional terms of service** duties require category 1 service providers to use proportionate systems and processes to ensure that taking down or restricting access to content, and suspending or banning users, is only carried out in accordance with their terms of service. Category 1 service providers must enforce consistently any provisions in their terms of service related to these actions. Ofcom will produce guidance for providers of category 1 services to assist them with complying with these duties.

Duties in the Act

3.1 In addition to the terms of service duties we consulted on under phase one[17] and the terms of service duties we will consult on under phase two[18], providers of services designated as category 1 will have additional duties relating to how they communicate and enforce certain provisions within their terms of service.

3.2 The additional terms of service duties only relate to provisions in the terms of service which indicate what content is prohibited on the service, to what content the provider will restrict user access, and in which cases the provider will suspend or ban a user from using the service.[19] In this section, we refer to the provisions to which these duties relate as "**relevant terms of service**".

3.3 Under the additional terms of service duties, providers of a category 1 service must **use proportionate systems and processes designed to ensure that they:**

a) **do not take down user-generated content, restrict users' access to user-generated content or suspend or ban users from using the service except in accordance with the relevant terms of service;**[20] and
b) **enforce any provisions related to taking down user-generated content, restricting users' access to user-generated content or suspending or banning users from using the service.**[21]

3.4 Providers of a category 1 service must also **ensure relevant terms of service are**:

a) **clear, easily accessible and communicated in sufficient detail** so that users can be reasonably certain whether the provider would be justified in taking the specified action in a particular case; and

[17] Service providers are also subject to various terms of service duties related to illegal content (sections 10(4)-(8), 21(3), 25(2), 27(5)-(8), 32(3) of the Act, see our illegal harms consultation). Providers of services publishing pornographic content are also subject to terms of service duties related to age verification or age estimation (section 81(5) of the Act – see our pornographic content providers consultation).
[18] Providers of services likely to be accessed by children are subject to various terms of service duties related to child safety (sections 12(9)-(13); 21(3); 25(2); 29(5)-(8); 32(3) of the Act – we will consult on these duties in our Spring 2024 protection of children consultation.
[19] Sections 71 and 72 of the Act.
[20] Section 71(1) of the Act.
[21] Section 72(3) of the Act.

b) **applied consistently**.[22]

3.5 Providers of a category 1 service **must provide easy to access, easy to use (including by children) and transparent reporting and complaints processes** that enable users and, where relevant, affected persons[23] to:

a) report content or users they believe are captured by the **relevant terms of service**;
b) complain about content they consider captured by the **relevant terms of service**;
c) complain where their content has been taken down or access to it has been restricted in relation to the **relevant terms of service**, or they have been banned or suspended from using the service; and
d) complain where they believe the service is not meeting its various duties under this section.[24]

Services **must also set out the complaints process in their terms of service and take appropriate action in response to complaints**.[25]

3.6 The additional terms of service duties do not prevent a provider from taking down content, restricting user access to content or suspending or banning a user where that action is taken:

a) to comply with its duties to protect individuals from illegal content;[26]
b) to comply with its duties to protect children from content that is harmful to them;[27]
c) to avoid criminal or civil liability on the part of the provider that might arise from not taking the action;[28]
d) on the basis that a user has committed an offence;[29] or
e) in relation to 'consumer content' (which is content that, among other things, amounts to an offer to sell goods or supply services), or the terms of service that deal with the treatment of consumer content.[30]

---

[22] Section 72(4) of the Act.
[23] An "affected person" is a person who is not a user of the service, who is in the United Kingdom and who is: the subject of the content; a member of a class or group of people with a characteristic targeted by the content; a parent of or other adult responsible for a child user who is the subject of the content; or an adult providing assistance to another adult who needs assistance to use the service and is either a user of the service or the subject of the content (section 74(6) of the Act).
[24] Section 72(5) to (9) of the Act. Category 1 service providers also have complaints duties under section 21 of the Act: they must operate a complaints procedure which enables users to complain that the provider is not complying with its duties related to user empowerment (section 15), content of democratic importance (section 17), news publisher content (section 18), journalistic content (section 19) and freedom of expression and privacy (section 22(4), (6) or (7)). There are also complaints duties for Category 1, Category 2A and Categories 2B services relating to the disclosure of information about use of a service by deceased child users (section 75(5), (6)).
[25] Section 72(6) and (7) of the Act
[26] Section 71(2)(a)(i).
[27] Section 71(2)(a)(ii).
[28] Section 71(2)(b).
[29] Section 71(3).
[30] Section 71(4).

Implementing the Act

3.7 Ofcom is required to produce guidance for providers of a category 1 service to assist them in complying with their duties in this area.[31]

Questions for stakeholders

3.8 In this section, we are interested in understanding more about how providers of online services currently or could:

a) ensure their terms of service enable users to understand when and how different types of enforcement action will be taken against content or accounts;
b) measure whether users read, understand, and follow their rules; and
c) avoid over- or underenforcing their terms of service when implementing moderation systems.

3.9 **We welcome responses on the questions below, particularly questions 3-5, 10-11, 12(c) and 13(b) which we consider key for the development of guidance in this area. We also welcome any additional evidence or information that stakeholders consider may be relevant.**

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.

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.

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) 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;

[31] Section 73 of the Act.

b) any relevant considerations about the risk of bad actors taking advantage of transparency around your terms of service and how they are enforced;
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; and
d) any other information.

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) how you measure user engagement with/comprehension of your terms of service and the metrics you collect;
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);
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.);
d) costs of these processes (including the design, implementation and continued use of these processes or updated versions of these processes);
e) how you evaluate the effectiveness of measures designed to improve engagement with and/or comprehension of your terms of service; and
f) any other information.

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:

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;
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);
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); and
d) any other information.

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, including:

a) reporting or complaints routes for registered users, non-registered users and potential complainants (being affected persons who are not users of the service);
b) how to ensure that reporting and complaints mechanisms are not misused;
c) the key choices and factors involved in designing these mechanisms;
d) how users can or should be supported to report/complain about specific concerns (e.g., other users, certain types of content or, appeal content takedowns or account bans);
e) how to ensure they are user-friendly and accessible to all users (e.g., disabled users, children);
f) whether users are informed that their reports are anonymous (e.g., other users will not be informed about who has reported their content or account);
g) any user support materials that explain how to use the reporting and complaints process and what will happen when users engage with these systems; and
h) any other information.

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 on:

a) how users report harmful content on your service(s) (including the mechanisms' location and prominence for users, and any screenshots you can provide);
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;
c) how users appeal against content takedowns, content restrictions or account suspensions or bans;
d) what type of content or conduct users and non-users may make a complaint about / report, including any specific lists or categories;
e) whether users need to create accounts to access reporting and complaints mechanisms (if there are multiple mechanisms, please provide information for each mechanism);
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;
    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;
    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;
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;
h) the use of trusted flaggers (and if reports from trusted flaggers should be prioritised over reports or complaints from users);

i) the cost involved in designing and maintaining reporting and/or complaints mechanisms, including any relevant issues, difficulties or considerations relating to scalability; and
j) any other information.

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) 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;
   ii) any differences for cases involving multiple reports/complaints about a single piece of content or user;
   iii) the costs associated with reviewing reports;
b) whether any reports or complaints are expedited or directed to specialist teams, including:
   a. the criteria for this; and
   b. the cost involved in facilitating this;
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;
   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); and
   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);
   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);
   v) the costs associated with responding to reports;
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);
e) any internal or external timeframes or key performance indicators (KPIs) for reviewing and/or acting on reports or complaints;
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; and
g) any other information.

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) improvements in terms of user safety and user rights (e.g., freedom of expression), as well as any relevant considerations around potential costs or cost drivers;
b) evidence of the effectiveness of existing moderation systems including any relevant examples of the accuracy, bias and or effectiveness of specific moderation processes; and
c) any other information.

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 include any relevant information relating to:

a) the types of action taken, and frequency of these actions (including per type of action);
b) how relevant content or users were or would be brought to your attention;
c) any policies, approaches or processes you have used or would use to guide moderation decisions in these cases;
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; and
    ii) when and how these new policies would be added to or included in your publicly available terms of service; and
e) any other information.

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 describe (with examples) any relevant circumstances that:

a) 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);
b) 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);
c) relate to competing priorities (e.g., freedom of expression, public interest concerns);

d) 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 exceptions for not removing violating content (e.g. news content), or transparency is not provided to avoid empowering bad actors); and
e) any other relevant information.

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 regarding:

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;
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;
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;
iv) how data is used to develop, train, test or operate content moderation systems is sourced for different provisions within your terms of service;
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);
vi) how and when automated systems are updated, and the trigger for this (e.g., in response to changing user behaviour or emerging harms); and
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;

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;
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; and
d) any other information.

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 regarding:

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;
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;
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;
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;

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; and
c) any other information.

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 regarding:

a) whether certain moderators are specialised in certain harms or subject material relating to different provisions in the terms of service;
b) how services can/should/do assess the accuracy and consistency of human moderation teams;
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);
d) whether training is provided and/or updated (including for emerging harms), and the frequency of these updates;
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);
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); and
g) any other information.

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 regarding:

a) how and when automated systems or human moderators are deployed in the moderation process;
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);
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;
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
e) any other information.

4. News publisher content, journalistic content and content of democratic importance

These duties place additional requirements on category 1 service providers to protect **news publisher content**, **journalistic content** and **content of democratic importance.**[32]

Specifically, category 1 service providers must not take action in relation to any **news publisher content**, or a user that is a recognised news publisher, before notifying the recognised news publisher in question. Ofcom will publish guidance on this duty to aid compliance.

For **journalistic content** and **content of democratic importance**, category 1 service providers are required to use proportionate systems and processes to ensure the importance of the free expression of this content is taken into account when moderating this content. Providers must include provisions in the terms of service specifying these systems and processes which are clear, accessible, and applied consistently. Providers must also have a dedicated and expedited complaints procedure in place in relation to journalistic content. Ofcom will produce codes of practice on these duties.

Duties in the Act

News Publisher Content

4.1 **News publisher content** is defined as content which is generated directly on a service by a "recognised news publisher" or is a reproduction, recording or link to an article or item originally published or broadcast by a recognised news publisher.[33] **A recognised news publisher**[34] is defined as the BBC, S4C, the holder of an Ofcom broadcast licence who publishes news-related material in connection with the broadcasting activities authorised under the licence or an entity which meets a number of specified criteria.[35]

[32] These are in addition to the duties placed on all service providers relating to freedom of expression and privacy. These duties require all in-scope service providers to have particular regard to the importance of protecting users' rights to freedom of expression and privacy when deciding on and implementing safety measures and policies (sections 22 and 33 of the Act).
[33] Sections 55(8) to (10) of the Act.
[34] A proscribed organisation under the Terrorism Act 2000 (or an entity whose purpose is to support a proscribed organisation) or a sanctioned entity under the Sanctions and Anti-Money Laundering Act 2018 is excluded from being a recognised news publisher. See sections 56(3) and (4) of the Act.
[35] For example, has as its principal purpose the publication of news-related material which is created by different persons and subject to editorial control; publishes such material in the course of a business (whether or not carried on with a view to profit); is subject to a standards code; has policies and procedures for handling and resolving complaints; has a registered office or other business address in the United Kingdom. See section 56(2) of the Act.

4.2 Category 1 service providers must **not take action**[36] **in relation to any news publisher content, or a user that is a recognised news publisher, before notifying the recognised news publisher in question**.[37] This notification must **specify the proposed action**, **give reasons by reference to relevant provisions** in the terms of service and provide the recognised news publisher with a **reasonable period to make representations.** If the content in question is also journalistic content (see below), the service provider must also **explain how it took the importance of the free expression of journalistic content into account**. Following consideration of any representations, the service provider must **notify the recognised news publisher of its decision and the reasons for it**, addressing any representations made.

4.3 If the service provider reasonably considers that it would incur criminal or civil liability for not removing this content in relation to the news publisher content or the content amounts to a relevant offence, it can take action first and then swiftly notify the recognised news publisher, specifying a reasonable period within which the recognised news publisher may request the action is reversed. If such a request is made, the service provider must **consider the request and whether the above steps should have been taken**; if so, it must **swiftly reverse the action and notify the recognised news publisher** of its decision and the reasons for it. Category 1 service providers do not have to comply with these duties if a recognised news publisher has been banned from using a service and the ban is still in force.

4.4 News publisher content is exempt from the safety duties imposed on user-to-user services, meaning service providers are under no legal obligation under the Act to take any action in relation to news publisher content under the illegal content and protection of children duties.

Journalistic Content

4.5 **Journalistic content** is defined as news publisher content or regulated user-generated content which is generated for the purposes of journalism and is UK-linked.[38] Content is UK-linked if UK users form one of the target markets for the content (or the only target market) or the content is or is likely to be of interest to a significant number of UK users.

4.6 Category 1 service providers must **operate a service using proportionate**[39] **systems and processes** designed to ensure that the importance of the free expression of journalistic content is taken into account when identifying such content, making decisions about how to treat such content and whether to take action[40] against a user generating, uploading or sharing such content.[41]

---

[36] Taking action in relation to content means taking down content, restricting users' access to content or adding warning labels to content, except warning labels normally encountered only be child users, and any other action in relation to content subject to a relevant term of service. See sections 18(13) and (14) of the Act.
[37] Section 18 of the Act. See section 18(9) for circumstances where a provider will not be regarded as taking action in relation to news publisher content.
[38] Section 19(10) of the Act.
[39] In determining what is proportionate, the size and capacity of the provider of a service, in particular, is relevant. See section 19(9) of the Act.
[40] Taking action against a user means giving a warning to a user, suspending or banning a user from using a service, or in any way restricting a user's ability to use a service. See section 19(12) of the Act.
[41] Section 19 of the Act.

4.7 Category 1 service providers must make a **dedicated and expedited complaints procedure** available in relation to journalistic content.[42] If a complaint is upheld, the content must be **swiftly reinstated on the service or the action against the user swiftly reversed.**

4.8 Category 1 service providers must also include **provisions in the terms of service** specifying by what methods content present on the service is to be identified as journalistic content; the policies and processes by which the importance of the free expression of journalistic content is taken into account; and the policies and processes for handling complaints in relation to content which is, or is considered to be, journalistic content. These must be **clear and accessible and applied consistently.**

Content of Democratic Importance

4.9 **Content of democratic importance** is defined as news publisher content or regulated user-generated content which is or appears to be specifically intended to contribute to democratic political debate in the UK or a part or area of the UK.[43]

4.10 Category 1 service providers must **operate a service using proportionate[44] systems and processes** designed to ensure that the importance of the free expression of content of democratic importance is taken into account when making decisions about (a) how to treat such content (especially decisions about whether to take it down or restrict users' access to it) and (b) whether to take action[45] against a user generating, uploading or sharing such content.[46] These systems and processes must be **applied in the same way to a wide diversity of political opinion.**

4.11 Category 1 service providers must include **provisions in the terms of service** specifying the policies and processes by which the importance of the free expression of content of democratic importance is taken into account. These must be **clear and accessible and applied consistently.**

4.12 While category 1 service providers are not required to have a dedicated and expedited complaints process in place for content of democratic importance, the complaints procedures duties do apply to this type of content.[47]

Impact Assessments

4.13 Category 1 service providers are also under additional duties to carry out various assessments regarding the impact of safety measures and policies on users' rights to freedom of expression and privacy. These must include **an assessment of the impact of safety measures and policies** on the availability and treatment of news publisher content and journalistic content in relation to the service.

---

[42] Category 1 services are not required to make a dedicated and expedited complaints procedure available to a recognised news publisher in relation to a decision if the provider has taken the steps set out in section 18(3) of the Act in relation to that decision.
[43] Section 17(7) of the Act.
[44] In determining what is proportionate, the size and capacity of the provider of a service, in particular, is relevant. See section 17(6) of the Act.
[45] Taking action against a user means giving a warning to a user, suspending or banning a user from using a service, or in any way restricting a user's ability to use a service. See section 17(8) of the Act.
[46] Section 17 of the Act.
[47] Section 21 of the Act.

4.14 Providers must **publish these impact assessments as well as the positive steps they have taken in response and keep impact assessments up to date**.[48]

Implementing the Act

4.15 The Act requires Ofcom to produce one or more codes of practice describing measures recommended for the purpose of compliance with the duties to protect journalistic content and content of democratic importance.[49] We must also produce guidance on the duties to protect news publisher content.[50]

Questions for stakeholders

4.16 In this section, we are interested in understanding more about:

a) the identification, classification, and moderation of the above content on in-scope services;
b) the effectiveness and cost implications of these processes; and
c) measures used to prevent the misuse of systems to identify and categorise this type of content and particularly, the current application of complaints and appeals processes.

4.17 The questions pertain to any content which might fall in scope of journalistic, news publisher or content of democratic importance as per the Act, even if it has been defined in a different manner by the online service provider in their terms of service.

4.18 **We welcome responses on the questions below. We also welcome any additional evidence or information that stakeholders consider may be relevant**.

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 particular, we are interested in:

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.

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.

[48] Section 22(4)-(7) of the Act.
[49] Section 41 of the Act.
[50] Section 52(2) of the Act.

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 particular, please explain:

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;
b) the factors that are or should be considered when taking action (e.g.: downranking/removal/suspension/ban or other) regarding this content;
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;
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; and
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.

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 particular, we are interested in:

a) examples of effective redress mechanisms that you consider would be most suited to these content types; and
b) briefings, investigations, transparency reports, media investigations and research papers that provide more evidence.

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 particular, please provide evidence of:

a) any known impacts of over-removal or erroneous removal of news publisher content, journalistic content or content of democratic importance; and
b) briefings, investigations, transparency reports, media investigations and research papers regarding misuse of such speech protective provisions.

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 particular, we are interested in:

a) any initiatives taken to create and increase awareness about the process for users to complain and/or appeal content removals; and
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.

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 particular, please:

a) explain which elements of your service design or operation they relate to and which factors they take into account; and
b) provide relevant briefings, investigations, transparency reports, media investigations and research papers.

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 particular, we are interested in:

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

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 particular, we are interested in:

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

5. User empowerment duties

The **user empowerment duties** require category 1 service providers to offer adult users features that can be applied to reduce their exposure to certain types of legal content, and filter out content from non-verified users. These tools must be easy to access. Ofcom will produce a code of practice, describing measures recommended for the purposes of complying with these duties, as well as guidance which contains examples of the content to which these duties apply. Category 1 service providers will also have to undertake a content assessment to, among other things, measure the incidence of this content on their service. Ofcom will produce guidance for the purposes of the assessment duties.

Duties in the Act

5.1 The Act places duties on category 1 service providers to provide features that empower adult users to have greater control over specific types of content they encounter online. For the purposes of this call for evidence, we refer to this content as **relevant content**, which is legal content that:

a) encourages, promotes or provides instructions for:

i) suicide or an act of deliberate self-injury; or
ii) an eating disorder or behaviours associated with an eating disorder;[51]

b) is abusive and targets race, religion, sex, sexual orientation, disability or gender reassignment;[52] or
c) incites hatred against people of a particular race, religion, sex or sexual orientation, people who have a disability, or people who have the characteristic of gender reassignment.[53]

5.2 The Act describes these categories of content in the same way as certain categories of primary priority content or priority content that is harmful to children.[54]

5.3 Category 1 service providers will have **assessment duties to**:

a) **undertake an assessment** for the purposes of the user empowerment duties[55] (including an assessment of the incidence of **relevant content** on their service)[56] and **keep this assessment up to date**;[57] and
b) **keep a written record of any such assessment**, including details about how the assessment was carried out and its findings,[58] and as soon as reasonably practical after making or revising such a record, **supply Ofcom with a copy of the record**.[59]

[51] Section 16(3) of the Act.
[52] Section 16(4) of the Act.
[53] Section 16(5) of the Act.
[54] Sections 61(3) and (4) and 62(2) and (3) of the Act.
[55] Section 14(2) of the Act.
[56] Section 14(5) of the Act.
[57] Section 14(3) and (4) of the Act.
[58] Section 23(9) of the Act.
[59] Section 23(10) of the Act.

5.4 Under the user empowerment duties, category 1 service providers must also:

a) to the extent that it is proportionate to do so, **offer all adult users control features that are easy to access** and that, if applied, result in the service using systems or processes designed to effectively:

i) reduce the likelihood of the user encountering relevant content; or
ii) alert the user to relevant content present on the service;[60]

b) use **systems or processes which seek to ensure that all registered adult users are offered at the earliest possible opportunity to indicate to the provider that they wish to turn these features on or off** (whether they are on or off by default);[61]
c) **specify in the terms of service which control features are offered and how users may take advantage of them**;[62]
d) include **a summary of the findings of their most recent assessment** in the terms of service;[63] and
e) offer adult users **features that, if used, prevent non-verified users from interacting with the users' content** and reduces the likelihood of the user encountering content from non-verified users.[64]

Implementing the Act

5.5 Ofcom is required to produce a code of practice[65] describing measures recommended for the purpose of compliance with the user empowerment duties, as well as guidance on the assessment duties[66] and guidance which contains examples of what Ofcom considers is or is not **relevant content**.[67]

Questions for stakeholders

5.6 In this section, for the purposes of the user empowerment code of practice, we are interested in understanding more about:

a) the detection, classification and moderation of relevant content on in-scope services; and

the tools and features that are currently offered to empower users to reduce exposure to relevant content (or other categories of permitted content which users may wish not to encounter) and interactions between different types of users, including the effectiveness, take-up and costs associated with them.

5.7 In this section, for the purposes of the assessment guidance we are interested in understanding more about:

a) the incidence of relevant content for adult users across in-scope services; and

[60] Section 15(3) and (4) of the Act.
[61] Section 15(5) of the Act.
[62] Section 15(7) of the Act.
[63] Section 15(8) of the Act.
[64] Section 15(9) and (10) of the Act.
[65] Section 41(3) of the Act.
[66] Section 52(1) and (3)(a) of the Act. [67] Section 53(2) of the Act.
[67] Section 53(2) of the Act.

b) the experience of adult users with a protected characteristic in encountering relevant content, or of those likely to be particularly affected by such content.

5.8 We have asked in our previous call for evidence for protection of children about the prevalence of **relevant content**, but here we are interested in how the impact of **relevant content** might be different for adults to children.

5.9 **We welcome responses on the questions below.[68] We also welcome any additional evidence or information that stakeholders consider may be relevant.**

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 particular, we are interested in:

a) what systems you use for detection;
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;
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;
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;
e) whether **relevant content** is allowed or prohibited on your service;
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; and
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.
h) If you have provided relevant information in response to complaints and reporting questions in Section 3, additional terms of service duties, you may cross refer to these here.

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.

[68] There is some cross-over with Ofcom's voluntary initiative - Best Practice Principles for Media Literacy by Design – which provide social media, gaming, pornography, sharing and search services of all sizes with guidance for how to approach media literacy on platform. Ofcom expects platforms to use effective tools to empower users and support them in providing context to the content they see.

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?

For providers of online services

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 particular, we are interested in:

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;
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; and
c) if you conduct any research into the experience of the above users on your service.

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?[69]

In particular, we are interested in:

a) features offered to users to reduce the likelihood of them encountering content they do not wish to see;
b) features offered to users to alert them to the presence of certain categories of content;
c) features offered to users to enable them to control their interactions with different types of users (e.g., non-verified);
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.**

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 particular, we are interested in:

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;
b) how you measure user engagement with these features, and any evidence you can provide around this;

[69] In recent Ofcom research we considered 'content controls' to be personalised settings provided by social media and video-sharing platforms, from Ofcom, 2023. Fewer than half of social media users find content controls to be effective.

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;
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) how you develop the user need for these features, and the factors considered when determining to develop them;
ii) whether these features are on by default, and in what circumstances;
iii) whether these features are personalised for specific types of users;
iv) when to offer users these features;
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);
vi) where users learn about these features;
vii) how to provide information about these features, including the level of detail and the words used to describe complex or technical concepts;
viii) whether users have choice of controls over specific types of content;
ix) how you decide whether to iterate, replace or keep such features; and
x) any other factors not already covered above that you take into account when considering such features; and
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).

6. User identity verification duties

The **user identity verification** duties require category 1 service providers to offer UK adult users the option to verify their identity. Ofcom will produce guidance to assist services in meeting these duties.

Duties in the Act

6.1 The Act places duties on category 1 service providers to provide UK adult users[70] with the option to verify their identity, where identity verification is not already required to access the service. Adult users will be able to control their exposure to all non-verified users (including users who do not need to be offered the option to verify their identity in accordance with this duty) via the user empowerment tools[71], set out in the section above.

6.2 The user verification duties require services to:

a) **offer all adult users the option to verify their identity;**[72] **and**
b) **explain how the verification process works in their terms of service in clear and accessible provisions**.[73]

6.3 The verification process may be of any kind and does not need to require documentation to be provided.[74]

Implementing the Act

6.4 The Act requires Ofcom to produce guidance for providers of category 1 services in order to meet the above duties, having particular regard to the desirability of ensuring verification is likely to be available to vulnerable adult users.[75]

Questions for stakeholders

6.5 In this section, we are interested in understanding more about:

a) the circumstances where identity verification is offered on user-to-user services, how this is done, and the cost and effectiveness of these methods; and

[70] This means an adult in the UK who (a) is a user of the service, or (b) seeks to begin to use the service (for example by setting up an account) (section 64(7) of the Act).
[71] The definition of "non-verified users" for the purposes of the user empowerment duties is a user who is an individual, whether in the UK or outside of it, who has not verified their identity to the service provider (section 16(7) of the Act).
[72] Section 64(1) of the Act.
[73] Section 64(3) of the Act.
[74] Section 64(2) of the Act.
[75] Section 65 of the Act.

b) the broader implications surrounding identity verification, including user attitudes towards verified accounts and user attitudes to widespread implementation (particularly as it relates to the experience of verifying the identity of individual adult users).

6.6 **We welcome responses to the questions below. We also welcome any additional evidence or information that stakeholders consider may be relevant.**

Circumstances where user identity verification is offered and how

For all respondents

Question 31: What kind of user-to-user[76] services currently deploy identity verification and in what circumstances? Including:

a) the ways in which these identity verification methods are beneficial, both to the user and to the service;
b) what documentation you understand to be necessary for different types, or levels, of identity verification on user-to-user services;
c) whether you believe there are there any other circumstances where identity verification should be offered on user-to-user services.

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;
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;
c) whether the process is, or can be, tailored to users in different geographical areas, such as the UK;
d) whether you engage third party providers to provide all or part of this identity verification process and, if so, which providers;
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;
f) how your identity verification policies have been developed, including any research that you can share;
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;
h) any consideration around users who may be vulnerable participating in the identity verification method;
i) how you manage the identity verification of users who have multiple accounts;

[76] As per the OSA, a user-to-user service means an internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, may be encountered by another user, or other users, of the service" (section 3(1) of the Act).

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[77], and how you consider user perceptions of these different methods;

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, including:

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;
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;
c) any other benefits or unintended consequences from these schemes existing; and
d) the safeguards necessary to ensure users' privacy is protected.

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

a) any unintended consequences of implementing identity verification, such as the impact this may have on your site's ecosystem;
b) how you envisage your service operating in the digital identity market, bearing in mind moves towards cross-industry and federated identity schemes.

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? Including:

a) whether they value verification being offered on a service;
b) whether verification influences user behaviour, such as whether they perceive identity verification to signify authenticity;

[77] We defined 'Monetised Scheme' in our illegal content consultation as "A scheme by which a service labels the user profile of a user who has made payment to the provider of the service or some other person. Such schemes may be open to all users and payment may be regular or one-off. Users participating in the scheme may benefit from access to additional features on the service. The label to indicate that a user is participating in a monetised scheme may appear on that user's profile and/or any content they publish. Services may or may not refer to such schemes as "verification" schemes," and 'Notable user scheme' as "a scheme by which a service labels the user profile of a user to indicate to other users that they are notable. "Notable users" include but are not limited to politicians, celebrities, influencers, financial advisors, company executives, journalists, government departments and institutions, nongovernmental organisations, financial institutions, media outlets, and companies. The label to indicate that a user is notable (for example a "tick" symbol) may appear on that user's user profile and/or any content they publish. Services may or may not refer to such schemes as "verification" schemes." From Ofcom, 2023. Annex 7 Illegal Content Codes of Practice for user-to-user services, pages 47 and 49.

c) attitudes towards non-verified, anonymous or pseudonymous users and the willingness to engage with them;
d) who you deem to be 'vulnerable' in terms of verifying their identity online – for example, whether this includes users unable to access or less likely to hold identification documentation, and those who may become vulnerable by displaying their identity to other users.

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? Including:

a) take-up of identity verification by your users;
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.

7. Fraudulent advertising

The **Fraudulent Advertising** duty requires providers of category 1 and 2A services to operate their services using proportionate systems and processes, designed to tackle fraudulent advertising. Ofcom will develop and publish a code of practice, describing measures recommended for the purpose of compliance with these duties.

Duties in the Act

7.1 The Act puts in place a special regime designed to ensure that category 1 and 2A services tackle fraudulent paid-for advertising. A 'fraudulent advertisement'[78] is a paid-for advertisement[79] which amounts to one or more of the specified fraud offences under the Act.[80] User-generated content is exempt from the definition of a fraudulent advertisement in relation to category 1 services[81], while paid-for advertisements are excluded from the definition of 'search content'.[82]

7.2 As part of their duties in relation to fraudulent advertising, category 1 and 2A service providers[83] will be required to **operate the service using proportionate systems and processes designed to**:

a) **prevent individuals from encountering content consisting of fraudulent advertisements by means of their service**;
b) **minimise the length of time for which any such content is present**;
c) where the service provider is alerted by a person to the presence of such content or becomes aware of it in any other way, **category 1 service providers must swiftly take down such content**, whereas **category 2A service providers must swiftly ensure that individuals are no longer able to encounter such content** in or via search results of the service.

7.3 Such services must also **include clear and accessible provisions in their terms of service** giving information about any proactive technology used by the service provider for the purpose of compliance with this duty.

7.4 In determining what is proportionate, services must **have particular regard to: (a) the nature and severity of potential harm** to individuals presented by different kinds of fraudulent advertisement, and (b) **the degree of control the provider has over the placement of the advertisements on the service**.[84]

[78] This definition should not be confused with the fraudulent representation of online advertisement impressions to generate revenue – commonly referred to as 'ad fraud'.
[79] The definition of a 'paid-for advertisement' is set out under section 236 of the Act.
[80] The offences which relate to the fraudulent advertising duties are set out within section 40 of the Act. User-to-user and search content that amount to a relevant fraud offence is in scope of the general safety duties of the Act; we have consulted on proposed Codes of Practice.
[81] Section 38(3)(c) of the Act.
[82] Section 57(2)(a)
[83] The duties are set out for category 1 and 2A services within sections 38 and 39 of the Act, respectively.
[84] Sections 38(5) and 39(5) of the Act

Implementing the Act

7.5 Ofcom is required to produce a fraudulent advertising code of practice, describing measures recommended for the purpose of compliance with these duties.[85]

Questions for stakeholders

7.6 In this section, we are interested in understanding more about:

a) the processes and mechanisms (including their effectiveness, the costs, and the risks of unintended effects) that in-scope services currently use to support both the delivery of advertising and detection of fraudulent advertising material;
b) suggested additional processes and mechanisms that could be implemented in order for relevant services to meet their duties in relation to fraudulent advertising; and
c) any relevant evidence regarding the role of third-party intermediaries involved in the process of serving ads on in-scope services and their relationship to those services.

7.7 Please note that where questions request information related to specific numbers or, such information is useful to us for the purpose of assessing the proportionality of potential code measures.

7.8 **We welcome responses to the questions below. We also welcome any additional evidence or information that stakeholders consider may be relevant.**

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[86]);
b) The harms associated with different kinds of fraudulent advertisements, the severity of such harms, and, if relevant, how this varies by user group;
c) The key challenges to successfully detecting different types of fraudulent paid-for advertising, and how these challenges can be minimised or resolved;
d) The prioritisation of suspected fraudulent advertising within all categories of harmful advertising queues, e.g. account verification, user reports, appeals; and
e) The proportion of fraudulent advertisements that are currently estimated to remain undetected by services' systems.

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?

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'.

[85] Section 41(4) of the Act.
[86] 'Malvertising' is a term used to describe cyberattack techniques involving the spreading malware via online advertising networks.

For providers of online services

Question 39: What proportion of all paid-for advertising on your service is identified as fraudulent advertising?

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?

Question 41: Please provide information regarding the proportion of successfully identified fraudulent advertisements that are identified via a) automated systems, b) human processes, c) user reports d) other (please provide further detail).

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).

Proactive technology

For all respondents

Question 43: Please provide any evidence you have regarding proactive technologies[87] 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;
b) A brief description of how these technologies are/could be integrated into the service;
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;
d) How proactive technologies are maintained and kept up to date;
e) Information related to the associated time and/or costs for set-up, operation, and human review;
f) The cost of integrating such technologies: (a) for the first time; and (b) when updating these technologies over time;
g) Whether there are cost savings associated with these technologies.

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 relevant information on 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;

[87] Proactive technology' consists of three types of technology: content identification technology, user profiling technology, and behaviour identification technology (subject to certain exceptions). It is defined in section 231 of the Act.

b) The role of (a) automated processing and (b) human processing in the verification process, and how they interact;
c) The costs associated with advertiser verification and how those costs vary as scale increases;
d) The percentage of advertiser accounts that are verified;
e) Whether advertisers are permitted to publish advertisements on the service while the verification process is ongoing;
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;
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.

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 relevant information on the following points:

a) The percentage of submitted advertisements which are reviewed both (a) prior to and (b) after publication;
b) The role (a) automated processing and (b) human processing play in the review process and how they interact;
c) The red flags which trigger advertisement review processes both (a) prior to and (b) after publication and the basis on which those red flags are selected;
d) The timescales for review;
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);
f) The costs associated with the review of submitted paid-for advertisements;
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?

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 relevant information on the following points:

a) The role of (a) automated processing and (b) human processing in the appeals process, and how they interact;
b) The level of proof required for an appeal to be accepted;
c) The most frequent bases for appeals against sanctions decisions on fraudulent advertising content;

d) The ratio of decisions that are appealed against;
e) The costs associated with appeals;
f) The proportion of appealed decisions which are upheld and overturned.

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 relevant information on 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;
b) What percentage of user reports of advertisements relate to suspected fraudulent content, and the processes for taking action in relation to such reports;
c) Any statistics you can share on (a) 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;
d) The criteria used to classify and prioritise user reports;
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;
f) Any measures taken to make user reporting tools accessible, easy to use and easy to find for users;
g) How transparency and communication is maintained with users who have submitted reports.

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.

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;
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;
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;
d) The accuracy of detection methods, including true positive and false positive rates;
e) The costs associated with the development/acquisition and deployment of these detection mechanisms;
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 a) via automated processes, b) by human moderators, c) by service users.

8. Access to information about a deceased child's use of a service

The **Deceased Child Users** provisions require providers of all categorised services to make clear what their policy is about disclosing information to the parents of a deceased child user about the child's use of the service. Service providers must have a mechanism for parents to easily find out what they need to do to obtain information and updates in those circumstances. Ofcom is required to produce guidance to support providers of categorised services in complying with these duties.

Duties in the Act

8.1 All categorised service providers are required to **set out their policy on dealing with requests from parents of a deceased child for information about the child's use of the service in the terms of service (or for category 2A services in a publicly available statement), and to provide mechanisms for parents to find out what they need to do to obtain information and updates in those circumstances**.[88] For the avoidance of any doubt, there are no duties in the Act on providers of categorised services to have a policy for sharing information with parents.

8.2 A child in this context is anyone under the age of 18.[89] A parent in this context could be any person with parental responsibility.[90]

8.3 Where providers of categorised services have a policy for disclosure on their service they must:

- **have a mechanism (a dedicated helpline or section of the service) by which parents can easily find out what they need to do to obtain information and updates** about the request in those circumstances;
- **include clear and accessible provisions in the terms of service**:
  a) specifying the procedure for parents of a deceased child to request information about the child's use of the service;
  b) specifying what evidence (if any) the provider will require about the parent's identity or relationship to the child; and
  c) giving sufficient detail to enable the child users and their parents to be reasonably certain about what kinds of information would be disclosed and how information would be disclosed.

[88] Section 75 of the Act.
[89] Section 236 of the Act.
[90] Section 75(10) of the Act.

- **respond in a timely manner to requests from parents of a deceased child** for information about the child's use of the service or for updates about the progress of such information requests;
- **operate a complaints procedure that**:

a) allows for complaints to be made by parents who consider that the provider is not complying with the relevant duties;[91]
a) provide for appropriate action to be taken in response to such complaints; and
b) is easy to access, easy to use and transparent.

8.4 Separately, Ofcom will have a discretionary power to require services to provide information about a child's use of a service where it is requested in connection with an investigation into the death of a child.[92] We will set out our policy for using our section 101 powers in due course.

Implementing the Act

8.5 Ofcom is required to produce guidance to assist categorised services in complying with their duties regarding deceased child users.[93]

Questions for stakeholders

8.6 In this call for evidence, we are interested in understanding more about:

a) what kinds of evidence services might require about the parent's identity or relationship to the child, and about the death of the child;
b) what kinds of information parents might request about their child's use of the service, what information services do or might provide and how, and the challenges or trade-offs of doing so;
c) how long it should reasonably take services to provide that kind of information; and
d) what mechanisms currently exist for parents to find out what they need to do to obtain information and updates in these circumstances, whether these are easy to use, and what other mechanisms might be made available.

We welcome responses to the questions below. We also welcome any additional evidence or information that stakeholders consider may be relevant.

Processes for requesting information about a deceased child's use of a service

For all respondents

In particular, we would be interested to hear from those that have requested information from services following the death of a family member, and organisations representing bereaved families:

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

[91] Sections75(1)-(4)
[92] Section 101 of the Act.
[93] Section 76 of the Act.

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

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?

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

For providers of online services

We are particularly interested to hear from providers of online services that have a policy about providing information about deceased users:

Question 54: What kinds of information do you provide and how do you provide this information?

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.

Question 55: How long does it typically take you to provide information in response to a request?

a) How long should it reasonably take services to provide information in these circumstances?

Complaints systems

For the following questions, if you have provided relevant information in response to complaints and reporting questions in Section 3, additional terms of service duties, you may cross refer to these responses here.

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?

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?

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?

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

A1. Responding to this call for evidence

How to respond

A1.1 Ofcom would like to receive responses by 5pm on 20 May 2024.

A1.2 You can download a response form from our website. You can return this by email or post to the address provided in the response form.

A1.3 If your response is a large file, or has supporting charts, tables or other data, please email it to os-cfe@ofcom.org.uk, as an attachment in **Microsoft Word format**, together with the cover sheet. This email address is for this consultation only and will not be valid after 20 May 2024.

A1.4 Responses may alternatively be posted to the address below, marked with the title of the consultation:

Online Safety Policy Delivery Team
Ofcom
Riverside House
2A Southwark Bridge Road
London SE1 9HA

A1.5 We welcome responses in formats other than print, for example an audio recording or a British Sign Language video. To respond in BSL:

- send us a recording of you signing your response. This should be no longer than 5 minutes. Suitable file formats are DVDs, wmv or QuickTime files; or
- upload a video of you signing your response directly to YouTube (or another hosting site) and send us the link.

A1.6 We will publish a transcript of any audio or video responses we receive (unless your response is confidential).

A1.7 We do not need a paper copy of your response as well as an electronic version. We will acknowledge receipt of a response submitted to us by email.

A1.8 **You do not have to answer all the questions in the call for evidence if you do not have a view; a short response on just one point is fine.** We also welcome joint responses.

A1.9 It would be helpful if your response could include direct answers to the questions asked in this call for evidence. It would also help if you could explain why you hold your views and provide supporting evidence.

A1.10 If you want to discuss the issues and questions raised in this document, please contact the Online Safety team by email at os-cfe@ofcom.org.uk.

Confidentiality

A1.11 Calls for evidence are more effective if we publish the responses before the call for evidence period closes. This can help people and organisations with limited resources or familiarity with the issues to respond in a more informed way. So, in the interests of transparency and good regulatory practice, and because we believe it is important that everyone who is interested in an issue can see other respondents' views, we usually publish responses on the Ofcom website at regular intervals during and after the call for evidence period.

A1.12 If you think your response should be kept confidential, please specify which part(s) this applies to and explain why. Please send any confidential sections as a separate annex. If you want your name, address, other contact details or job title to remain confidential, please provide them only in the cover sheet, so that we don't have to edit your response.

A1.13 **If someone asks us to keep part or all of a response confidential, we will treat this request seriously and try to respect it, either by not publishing the response at all, or by only publishing the bits that are not confidential. Sometimes we might think it is important to disclose parts of a response that have been marked as confidential for reasons of transparency, but we will consult you before we do.** Occasionally we might have a legal obligation to publish information or disclose it in court, but again, as far as possible, we will let you know.

A1.14 Even if your response is not marked as confidential, we might still decide not to publish all or part of it in certain circumstances. For example, if we have concerns about the impact on your privacy or the privacy of others, that the content of the response might facilitate the commission of crime, or about the sensitive nature of the content more generally. If we decide not to publish all or part of your response, we will still take it into account in our consideration of the matter.

A1.15 To fulfil our pre-disclosure duty, we may share a copy of your non-confidential response with the relevant government department before we publish it on our website.

A1.16 Please also note that copyright and all other intellectual property in responses will be assumed to be licensed to Ofcom to use. Ofcom's intellectual property rights are explained further in our Terms of Use. Please also see our Privacy Statement.

Next steps

A1.17 If you wish, you can register to receive mail updates alerting you to new Ofcom publications.

Ofcom's processes

A1.18 Ofcom aims to make responding to a call for evidence as easy as possible.

A1.19 If you have any comments or suggestions on how we manage our calls for evidence or consultations, please email us at consult@ofcom.org.uk. We particularly welcome ideas on how Ofcom could more effectively seek the views of groups or individuals, such as small businesses and individual users of online services, who are less likely to give their opinions through a formal consultation.

A1.20 If you would like to discuss these issues, or Ofcom's consultation processes more generally, please contact the Corporation Secretary:

Corporation Secretary
Ofcom
Riverside House
2a Southwark Bridge Road
London SE1 9HA
Email: corporationsecretary@ofcom.org.uk

A2. Call for evidence coversheet

Basic details

Call for evidence title:

To (Ofcom contact):

Name of respondent:

Representing (self or organisation/s):

Address (if not received by email):

Confidentiality

Please tick below what part of your response you consider is confidential, giving your reasons why

- Nothing ☐
- Name/contact details/job title ☐
- Whole response ☐
- Organisation ☐
- Part of the response ☐

If you selected 'Part of the response', please specify which parts:

If you want part of your response, your name or your organisation not to be published, can Ofcom still publish a reference to the contents of your response (including, for any confidential parts, a general summary that does not disclose the specific information or enable you to be identified)?

Yes ☐ No ☐

Declaration

I confirm that the correspondence supplied with this cover sheet is a formal call for evidence response that Ofcom can publish subject to the confidentiality section above. However, in supplying this response, I understand that Ofcom may need to disclose some information marked as confidential where it is proportionate and fair to do so to enable appropriate consultation, or if Ofcom is ordered to disclose them. If I have sent my response by email, Ofcom can disregard any standard e-mail text about not disclosing email contents and attachments.

Ofcom aims to publish responses at regular intervals during and after the call for evidence period. If your response is non-confidential (in whole or in part), and you would prefer us to publish your response only once the consultation has ended, please tick here.

Name Signed (if hard copy)