ONLINE SAFETY ACT NETWORK

CHILD SAFETY DUTIES

The second phase of Ofcom's major consultation programme to inform their implementation of the Online Safety Act focuses on the duties to protect children. In this explainer, Prof Lorna Woods and Dr Alexandros Antoniou set out what these duties are and what they require from regulated services in response.

Overview

User-to-user services and search services are subject to duties in relation to content harmful to children under the Online Safety Act (OSA), but only when those services "are likely to be accessed by children" (s 7(4) OSA, repeated in ss 11(1), 12(1), 28(1) and 29(1); see also Explanatory Notes (EN), paras 222 and 227). All providers of a Part 3 service must carry out a "children's access assessment" (ss 35-36) to determine whether they are services to which children's safety duties apply. If so, user-to-user and search services must comply with the relevant risk assessment and safety duties. These obligations depend on Ofcom producing guidance and codes of practice (s 41). The consultation on these codes and guidance is due to be published by Ofcom on 8 May.

When is a service "likely to be accessed by children"?

The OSA puts the obligation on service providers to check whether they are a service to which the children's duties apply. Service providers must carry out an assessment within a timeframe specified in the OSA (keeping a record thereof). This is the children's access assessment. The assessment must be "suitable and sufficient" (s 36(6)). If the provider does not carry out an assessment or does not carry out an adequate one, that provider will be treated as though it were a service to which the duties apply (s 37(4) and (6)). Ofcom can give a confirmation decision (which is part of the enforcement process set out in the OSA) to a provider if it is satisfied that they have failed to comply with such duties (s 135(4) and (5)).

The aim of the children's access assessment is to check whether, if it is possible for a child to access the service, the "child user condition" is met. This is essentially the test for whether the service will be subject to the children's safety duties. The OSA specifies that a provider may only

"conclude that it is not possible for children to access a service, or part of it, if age verification or age estimation is used" (s 35(2)). The definition of age verification/age estimation excludes processes which rely on self-declaration of age (s 230(4)). Note that as regards the children's access assessment in ss 35 and 36, there is no equivalent of s 12(6), which requires for the purposes of the user-to-user safety duty that age verification be "highly effective" at correctly determining age. Nonetheless, s 35(2) states that the age verification measures should have the consequence that "children are not normally able to access the service or that part of it".

If this exclusory requirement is not met, the question then is whether the "child user condition" is met (s 35(3)) – either with regards to the whole service or part of it. This test can be satisfied in one of two ways:

- *either* there is a significant number of children who are users of the service/part of the service;
- *or* the service/part of service is of a kind likely to attract a significant number of users who are children.

For the first test, the intentions/target users of the service are irrelevant. The question is whether the threshold of "significant number" is met by reference to the number of actual users. The assessment should be repeated regularly (s 36(3)). "Significant" for these purposes is defined as "significant in proportion to the total number of United Kingdom users of a service" (s 35(4)(a)), but no further detail is included in the Act on when this threshold would be deemed to be passed. It is to be hoped that there is some similarity between Ofcom's approach and that of the ICO under the Children's Code. Indeed, the two regulators issued a joint statement in 2022 setting out how they proposed to "maximise coherence" between their respective online safety and data protection responsibilities and an updated Memorandum of Understanding has been promised now that Ofcom has received its powers under the OSA.

Children's Duties

As with the illegal content duties (see our Explainer here), the duties here are divided so that there are distinctions between:

- duties imposed on user-to-user services and search services;
- content harmful to children, priority and primary priority content; as well as
- specific risk assessment and risk mitigation duties.

Similarly, there are ancillary duties requiring systems to allow content reporting and complaints; and "cross-cutting duties" in relation to freedom of expression and privacy as well as record-keeping and review duties.

Meaning of Content Harmful to Children

Content harmful to children falls into one of three categories (s 60):

- primary priority content (s 60(2)(a));
- priority content (s 60(2)(b)); and
- other content "of a kind which presents a material risk of significant harm to an appreciable number of children in the United Kingdom" (s 60(2)(c)) – this is called "non-designated content that is harmful to children" (s 60(4)).

Essentially, non-designated content is a fall-back category for content that is not listed as either primary priority or priority content. While early versions of the Bill specified that illegal content would not constitute content harmful to children, it seems this limitation has been removed. Content where the risk of harm arises from the content's financial impact, or the safety or quality of goods featured in the content, or the way a service is performed does not fall within the definition of content harmful to children (s 60(3)).

There are four types of primary priority content (s 61; EN, para 366), as follows:

- pornographic content;
- content encouraging, promoting or providing instructions for suicide;
- content encouraging, promoting or providing instructions for deliberate self-injury (including poisoning);
- content encouraging, promoting or providing instructions for an eating disorder or behaviours associated with an eating disorder.

Note that pornographic content here relates to user-generated pornographic content (indeed all the content here is user-generated); what we might term commercial pornography is dealt with in Part 5 of the OSA.

There are six categories of content that are priority content harmful to children (s 62; EN, para 368):

- bullying content (and bullying is defined at s 62(12));
- content that is abusive or incites hatred based on any of race, religion, sex, sexual orientation, disability or gender reassignment;
- content encouraging, promoting or providing instructions for an act of serious violence against a person;
- content depicting real or realistic violence against a person or against an animal, or depicting real or realistic serious injury of a person or an animal in graphic detail (including fictional creatures);

- content encouraging, promoting or providing instructions for a challenge or stunt highly likely to result in serious injury;
- content encouraging a person to self-administer a physically harmful substance.

This is a slightly different list from that originally proposed by the Government at the first stage of Commons Report (Written Statement made by the then Secretary of State, Nadine Dorries, 7 July 2022 (UIN HCWS194)): the reference to harmful health content (including vaccine misinformation) was not included. Presumably, this could be caught by the category of non-designated content harmful to children.

The baseline definition of content harmful to children contains a threefold test of materiality, significance and appreciability, namely:

- a material risk;
- of significant (physical or psychological) harm (s 234(1)) such as serious anxiety and fear, depression and stress and other medically recognised mental illnesses (EN, para 958), whether this be because of the nature of the content (e.g., grossly offensive, abusive or discriminatory content, the manner of its dissemination (many people repeatedly sending content to an individual, or the fact of its dissemination (e.g., malicious sharing of personal information) and includes indirect harm (EN, paras 959-961); and
- to an appreciable number of children in the United Kingdom (and it is an open question as to whether this under-protects those in minoritised groups).

Risk Assessment Duties

Both user-to-user and search services must carry out a "suitable and sufficient" risk assessment and keep it up to date (s 11(2) and (3)); Ofcom is to provide guidance on risk assessments. The focus is on primary priority and priority content, with each type of content separately identified and bearing in mind the different age groups. Where a user-to user-service (but not search) identifies non-designated content harmful to children, it must notify Ofcom of that (s 11(5)). Services must take into account the extent to which the design of the service and its functionalities affect the level of risk (s 11(6)), highlighting specifically in relation to user-to-user (but not search) functionalities that enable adults to search for other users of the service (including children) and enabling adults to contact those other users (though this might be seen to be also relevant to some of the criminal offences caught by the OSA).

Risk Mitigation Duties ("Safety Duties")

The safety duties for user-to-user services and search services are different, with greater obligations falling on the former.

User-to-user services have a baseline duty to "take or use proportionate measures relating to the design or operation of the service" to mitigate and manage risks of harm, taking into account the risk assessment (s 12(2)(a)). Only the non-designated content identified by the risk assessment need be dealt with in respect of this general duty (s 13(2)). Further, there is a duty to design a system so as to protect children in age groups at risk of harm from certain types of harmful content present on the service (s 12(2)(b)). This could include signposting child users to sources of support if they have experienced harm (EN, para 109), so not actually requiring the prevention of harm.

Additionally, there is a specific duty in relation to primary priority content to operate the service in a way designed to prevent children of *any* age from encountering such content, with a requirement that this involves age verification/age estimation (s 12(3)(a) and (4)). Any such techniques must be highly effective at determining the correct age of the person. As for the illegal content duty (s 10(2)(a) and (4)), compliance with this obligation should look to the design of the system rather than whether any children have actually seen primary priority content.

This particular duty does not apply to harm arising from the dissemination of content but only the content itself. The EN give the example of doxing (para 121), where the information is not inherently problematic but the fact that it has been released is. It is unclear how the *manner* of dissemination is to be treated (as opposed to the *fact* of its dissemination). The other duties extend to harm from the fact or manner of dissemination (e.g., cumulative impact – see s 234(4)(a), and the EN make clear that this can be through operation of algorithms and other functionalities (para 960) as well as the content.

Search services also have general mitigation duties. The specific duty in relation to primary priority harm is not however to operate a system designed to prevent but rather to operate a system designed to minimise the risk of children of any age encountering relevant content (s 29(3)). The EN (para 196) suggest that this could include down-ranking, ensuring predictive searches do not drive child users towards harmful content and signposting towards resources and support.

When determining what is "proportionate" for the purposes of compliance with the child protection requirements, the findings of any children's risk assessment that has been conducted – including the likelihood and the severity of harm - and the size and capacity of the service provider will be relevant factors.

Ancillary Duties

Service providers (s 20 for user-to-user and s 31 for search) are required to operate a service with systems and processes that allow users and "affected persons" to easily report content

which they consider to be content that is harmful to children as well as providing an easy to access and use (including easy to use for children) complaints procedure (s 21 for user-to-user and s 32 for search), allowing complaints about content that is harmful to children.

There are also transparency obligations (Part 4, Ch 5) in relation to informing users as to how these obligations are to be met. So, user-to-user services must specify in the terms and conditions (or another form of a publicly available statement) how the above and other protections under the OSA are to be achieved, including the use of any proactive technology. There are, in addition, general record keeping and review obligations that apply (ss 23 and 34 for user-to-user and search services respectively).

April 2024