The Safety Net – Protecting children online

Safety first in the online playground

There is little doubt that children benefit from many online services, in particular from online learning, entertainment, and from a limited degree of online social interaction. Despite these positives, the harms of the internet and the harms identified in our article on the subject pose significant risks to children, with the potential to severely impact their mental and physical wellbeing. A primary goal of the Online Safety Act (OSA) is therefore to safeguard children in the digital world. To achieve this, the OSA imposes specific 'child safety obligations' on in-scope services (see our article on the scope of the act) which require them to be built with safety in mind, ensuring that they are not only safe by design but also offer a *higher* level of protection for children compared to adults.

So, how can online services determine whether they are within scope of these child safety obligations, exactly which obligations apply to them, and how to comply with them? Ofcom has published the following documents which can help to answer these questions:

- Child Safety Code for user-to-user services
- Child Safety Code for search services
- guidance on Children's Access Assessments
- guidance on Children's Risk Assessments
- guidance on Highly Effective Age Assurance

These documents are lengthy and contain quite a granular level of detail. To help break all this down, we have simplified and summarised Ofcom's key recommendations below by separating them into three compliance phases.

Phase One: Assess whether children are likely to access your service

All in-scope services (irrespective of whether they are U2U or search services must assess whether their service (or part of it) is likely to be accessed by children. To help with this assessment, Ofcom's guidance on Children's Access Assessments explains that, essentially, carrying out a Children's Access Assessment ('CAA') is a two-stage process which requires the service provider to determine:

1. if it is *possible* for children to access the service; and
2. whether there are significant numbers of children using the service or if the service is likely to attract a significant number of children.

On the first limb of the test, a service provider can *only* conclude that it is not possible for a child to access their service if they are using 'highly effective' age assurance (essentially an 'over 18s only' age-gating mechanism), which prevents children from accessing the service. In practice, the 'highly effective' threshold will be challenging to meet. The Guidance on Highly Effective Age Assurance explains that 'highly effective' means technically accurate, robust, reliable, and fair, emphasising that photo ID matching, facial age estimation, and reusable digital identity services would *potentially* meet this principles-based threshold, but that self-declaration of age would not. Notably, Ofcom has not proposed the use of numerical thresholds to establish whether an age-verification method is sufficiently 'highly effective'.

If a service *can* confidently reach an assessment that the service cannot be accessed by a child, the OSA child safety obligations will not apply. If children *can* access the service, the provider should then consider the second limb of the test: whether a significant number of children are likely to use the service or if the service is likely to attract a significant number of children. If a service does not have effective age assurance in place, it will likely be challenging for a provider to distinguish between adult and child users (and therefore to determine that a significant number of children use the service). Assessing whether the service is *attractive* to children might be easier - key indicators will be whether it is appealing visually, whether children form part of the commercial strategy, and whether the service benefits children.

The outcome of the CAA (and evidence used to reach the relevant outcome) should be recorded and continuously reviewed as part of the organisation's overall governance and accountability framework. The deadline for inscope service providers to complete their CAAs was 16 April 2025.

Phase Two: Assess the risks of the service

If, following Phase One, a service is considered 'likely to be accessed by a child', the next phase will be to complete a 'Children's Risk Assessment' (CRA). Ofcom has helpfully published Children's Risk Assessment Guidance to assist service providers with this exercise. In its guidance Ofcom ultimately proposes that a four-step methodology is adopted as part of this Phase:

1. Step One: Familiarise yourself and understand content that could be harmful.
2. Step Two: Assess the likelihood of children encountering each harm and assign a risk level for each kind of content harmful to children.
3. Step Three: Implement safety measures.
4. Step Four: Report, review, and update the CRA.

We consider steps three and four under Phase Three below. With respect to steps 1 and 2 and *understanding* online risks, Ofcom provides examples of content that could be harmful along with the corresponding risk factor and level associated with each harm in its draft Children's Risk Assessment Guidance (see Annex A1). Specific risk factors for U2U services include whether the service has a young user base, offers messaging functions, offers live streaming, uses recommender systems, or permits rapid forwarding / re-posting of content. Risk factors for search services include whether the service is a general or vertical search service, whether there is a predictive search function in the service, whether the service offers video or image searching, and the age or demographic profile of the users.

Ofcom has also published a draft 'Register of Risks' (see section 7 of the Consultation on Content Harmful to Children) which groups harms into the below categories. This register, which provides detailed evidence on risk factors, must be consulted by in-scope service providers when completing CRAs.

'Primary Priority Content' that is harmful to children	This is considered the most serious risk category and covers pornographic content, content relating to or that promotes or encourages suicide and self harm, and eating disorder content.
'Priority Content' that is harmful to children	This covers abuse and hate content which targets religion, sex, race, sexual orientation, disability, or gender reassignment. This also covers content which relates to or encourages bullying, violence, or harmful substances.
Non designated content	This covers content which is not Primary Priority or Priority Content of a kind which presents a material

	risk of significant harm to an appreciable number of children in the United Kingdom. Potential examples provided by Ofcom include 'body image content' and 'depressive content'.

As with the CAA, service providers should keep a written record of the CRA. This is separate to the Illegal Content Risk Assessment - see our article on illegal harms - that must also be carried out under the OSA.

The finalised Codes largely build on the measures outlined in the draft Codes and confirm the previous 4-step structure for the children's risk assessments.

Phase Three: Adopt safety measures to protect children

If the service is likely to be accessed by children and presents risks to children, it will be necessary to use proportionate measures to ensure the safety of the service. Again, Ofcom's specific draft guidance on mitigating risks is helpful and proposes more than 40 safety measures to help U2U and search services comply with their obligations. In many cases, the appropriate measure to be adopted may depend on the nature of the service (U2U or search), the size of the service, and the severity of the potential harms that could be encountered on the service. We have briefly summarised the key recommendations from Ofcom's draft codes of practice (see here for U2U services, see here for search services) in the table below.

Measure	**Description & Practical Compliance Tips**
Robust age checks	Adopting highly effective age assurance will work towards mitigating the risks for children on the service and may bring the service outside the scope of the OSA. Ofcom suggests that all U2U services that host or disseminate Primary Priority Content should implement highly effective age assurance. It may be the case that such age assurance will apply to the service in its entirety or to part of the service.
Safer algorithms	Personalisation has always been considered to present risks to children and can be a potential pathway to online harm. Services that use recommender systems should configure their algorithms to *filter out* the most harmful content. Ofcom also suggests that U2U services operating recommender systems *should not* recommend any Primary Priority Content to children. Amendments to the Codes now give providers a choice whether to *exclude illegal priority content from children's feeds* (as opposed to lowering its degree of prominence as was originally suggested), to suggest more *age differentiated online experiences.* This recognises the evolving capacities of children as they age and does not create a blanket measure to protect all children, irrespective of their age.
Effective moderation	U2U and search services should restrict content that is harmful to children through effective moderation. This moderation can be done automatically, by way of human moderation or through a combination of the two. Large search services should deploy a 'safe search' setting where, if a user is believed to be a child, Primary Priority Content in particular should be identified, downranked, and if necessary, blurred out.
Strong governance & accountability	A strong governance and accountability framework will be essential to ensure continued compliance with the child safety obligations of the OSA. This means that there should be adequate oversight over decision making, allocated roles

	and responsibilities, and effective reporting and review mechanisms. Ofcom suggests naming an individual with responsibility for online child safety, carrying out annual senior-body reviews, and implementing a Code of Conduct with standards for protecting children online.
More choice and support for children	It will be necessary to give children more information and control over their online experience. 'More information' means that U2U and search services should have clear terms and statements regarding the protection of children and should also make available the key findings of the relevant CRA. To provide children with 'more control', Ofcom recommends better explaining the complaints process, acknowledging receipt of complaints, explaining next steps, and offering an easy way to report predictive harmful search suggestions. U2U services should also build in controls to allow children to accept or decline group chat invites, block user accounts, and disable comments on their posts.

The above are *recommended measures only* and service providers can choose to comply with the child safety obligations using alternative measures. However, Ofcom makes it clear that services that choose to implement the measures set out in its codes will benefit from 'safe harbour treatment'. This means that they will be treated as complying with the child safety duties under the OSA and that Ofcom will not take enforcement action against them in relation to these duties.

Conclusion & Next Steps

Although the Codes are subject to final parliamentary approval, they are not likely to change at this point. Accordingly, if they have not done so already, service providers should think about: (1) how the OSA impacts them; (2) how they might be categorised under the OSA and what obligations might apply to them; (3) what child safety risks might be present on their service; and (4) what structures should be implemented to carry out the required assessments and adopt the mitigating measures recommended by the codes. This will enable service providers to have a strong foundation in place on which they can build the blocks of their OSA child safety compliance.

An in-scope provider which has assessed its service is "likely to be accessed by children" now has until **24 July 2025** to complete its children's risk assessment. Importantly, Ofcom may request this, so businesses should be able to justify the accuracy of the information in their assessment.

All this means that, from **25 July 2025**, in-scope providers should apply the risk mitigation measures set out in the Codes or alternative measures which are equally effective, noting that Ofcom will then be able to enforce against providers which fail to comply.

Last updated: 30 May 2025