The Online Safety Act: Children's Duties, Age Verification and Content Moderation on User-to-User Services

---

This explainer covers some of the issues arising following the implementation of the Online Safety Act children's duties, particularly in relation to age verification and the impact on users' access to content. It also addresses some of the misconceptions about the Act's requirements for content moderation and takedown that have been circulating since the duties came into force on 25 July 2025.

For more detail on the Online Safety Act children's duties, please see our website explainers on how they work and what their coming into force means.

THE CHILDREN'S CODES AND HIGHLY EFFECTIVE AGE ASSURANCE

Is age verification required?

The Online Safety Act (OSA) requires services likely to be accessed by children (following a children's access assessment under s 35) to do a children's risk assessment (s 11) and then to take steps to mitigate (s 12). The Act specifically requires that children should be prevented from encountering primary priority content (s 61 – pornography, suicide and self-harm material, eating disorder content) – and this is via highly effective age assurance (HEAA) (s 12(3)(a) and s 12(6)). HEAA includes either age verification or age estimation techniques.

So if a service doesn't have a risk of primary priority content, then age verification is not necessarily required (at content or service level) by the terms of the Act. Ofcom's codes contain further elaboration on when HEAA should be used and the question of whether there should be content or service-level age verification.

What is highly effective age assurance (HEAA)?

The OSA has minimum standards for age assurance and requires Ofcom to publish guidance on this (s 82), which was consulted on last year and published in its final form on 24 April 2025.

Ofcom has set down criteria in its guidance by which it assesses whether age assurance is highly effective: technically accurate, robust, reliable and fair. It also notes that any age assurance must comply

with the data protection requirements set down by the ICO. Ofcom suggests a non-exhaustive list of possible mechanisms, though it is for the service provider to choose the most appropriate method for its service: open banking, photo ID matching, facial age estimation, mobile network operator age checks, credit card checks, digital identity services and email-based age estimation.

Ofcom has launched an enforcement programme around age assurance. If age verification or age estimation set ups are easily tricked then arguably this is not highly effective.

There is a difference in enforcement context between Ofcom starting to prove that services have no HEAA in place, which can be obviously also checked by any user attempting to gain access to the service, as opposed to whether a service has deployed ineffective Age Assurance (which requires more evidence).

What content is covered by the children's safety duties?

As noted above, HEAA is required in relation to primary priority content (so news reporting is not caught). Priority content (s 62 – a longer list than primary priority) is content that is harmful to children which is to be mitigated but not necessarily by HEAA. Ofcom has included in its Children's Content code the possibility that HEAA could be used to target other measures aimed at protecting children (see below). Services can choose an editorial policy to go further in restricting access to this content if they choose, but that is their choice.

If a company uses HEAA to prevent access to a wide range of content that is not primary priority content on the basis of allegedly complying with its OSA duties, there is a question as to whether the company is satisfying its duties with regard to freedom of expression (s 22). There is less detail/profile on this aspect to date, though the rights of users to complain and requirements about clarity of terms of service support the freedom of expression duty.

What do Ofcom's children's codes say about HEAA?

Ofcom's codes provide more detail on when HEAA could be used beyond the circumstances required by the OSA. In particular, if the principal purpose of a service is to host or disseminate a kind of primary priority content (PPC) that is harmful to children, such as pornography, or priority content that is harmful to children, such as violent content, then the service should age-gate the entire service. So this measure isn't just about PPC but covers priority content in certain circumstances. Additionally, HEAA can be used to target appropriate safety measures at children, eg limiting the prominence of bullying content (which is priority not PPC). Using HEAA here means adults can still access that content as previously.

Does the Act require ID checks and facial scans?

The Act does not require ID checks. It anticipates age assurance (age verification or age estimation) on services likely to be accessed by children in relation to certain sorts of content (eg pornography). A service accessible by children which did not contain content harmful to children (or where the risk was really low) might not need to use age verification. Age assurance must be highly effective but the Act does not specify any particular technology (see above).

Does the Act prohibit the use of VPNs?

No, and Ofcom's Chief Executive, Melanie Dawes, in a Parliamentary hearing a few months ago, was clear that it was not possible under the Act to prevent their use. While the spike in downloads of VPNs has garnered lots of headlines in recent days, this is not surprising and has happened in other jurisdictions where age verification measures have been introduced. It's important to remember that the vast majority of children will not be downloading a VPN to access online content: the age assurance measures that are now in force will, however, protect those children, particularly the youngest, from involuntarily or otherwise accessing pornographic or violent content on social media, which was the purpose of these measures in the Act. Indeed, this is content that the vast majority of children themselves say that they do not want to see.

CONTENT MODERATION AND TAKEDOWN

Does the OSA create a general duty to remove protest content?

No, there is no general duty to remove content but the Act expects content moderation systems to be in place. There are limitations on the scope of these obligations because only certain types of content may trigger the illegal content and/or children's safety duties.

The Act creates duties to risk assess and mitigate harms arising from two groups of content – "illegal content" and "content harmful to children". To trigger any duties, content must fall into one of these two categories so this would depend on the nature of the "protest" videos. News reporting would not fall within this category; videos inciting violence against immigrants and containing racial hatred could; very violent content might trigger some of the duties relating to content harmful to children.

The Act contains no requirement that enables the regulator or the government to identify specific items of content and have those items of content taken down. In relation to certain types of illegal content identified in the Act (see Schedules 5, 6 and 7), services are under an obligation to "prevent individuals from encountering" it. The most obvious way to do this is by moderation and removal of content though *this is not expressly required by the Act*.

The Act requires services to have a system to take down illegal content once they have been notified of it. Since the duties relate to having a system in place to deal with the issue that is proportionate to the problem, a 100% success rate of the system is not required; service do not have to proactively search for relevant content. Ofcom in its Illegal Content Code for social media services (ICU Code) does require the removal of accounts of proscribed organisations (under terrorism laws) (see ICU Code, measure H1).

There is no obligation under the children's duties (which apply to a broader range of content) to take content down – the obligation is to prevent children from being harmed by the content.

When services take steps to comply with these duties they must also "have particular regard to the importance of protecting users' right to freedom of expression within the law" (s 22(2)).

What are the duties to protect content of democratic importance and journalistic content?

These duties will only apply in relation to a subset of user-to-user services designated as category 1 and will only come into force once the rules relating to the categorisation process have been finalised. This process has been delayed by the Wikipedia judicial review action. As such, there is no guidance from Ofcom yet on this. Ofcom's approach to moderation workflow under the ICU Code was to leave choices about priorities to the service provider.

In edge cases (protest livestreams that also capture crime or violence), how should platforms handle such mixed content under the Act?

The regime is not set up to consider edge-cases individually but how the service provider deals with content decisions in general – so the regime is more concerned with clarity of terms of service, training of moderators and the like. Ofcom has also produced guidance on how to approach questions of whether content is illegal content or not – the Illegal Content Judgments Guidance - and has also given examples of content types that in its view fall in or outside the children's content regime. Services can still enforce their own standards, which might also cover some of the edge case situations. *Note Ofcom is consulting on safeguards around livestreaming at the moment, with a focus on grooming of children and the livestreaming of terrorist attacks (eg Christchurch).*

What due-process obligations (appeals, transparency to users) apply when content is removed or restricted?

The Act requires service providers to have a complaints process (s 21) and the process must be detailed in the terms of service. This applies to all the content duties and the duty to have regard to freedom of expression. Ofcom has elaborated on this requirement in its Code. For example, in the ICU Code it specifies that not only must the complaints process exist it must be easy to find and easy to use, and that the services should inform complainants about how the complaint will be handled, giving time frames. Service providers are to determine appeals promptly. Ofcom also envisages that large services should monitor their performance as to the time it takes to determine an appeal and the accuracy of decision making.