BRIEFING: PETITION DEBATE ONLINE SAFETY ACT

December 2025 James Baker, Jim Killock

Table of Contents

Open Rights Group (ORG) is a UK based digital campaigning organisation working to protect fundamental rights to privacy and free speech online. We are a grassroots organisation with supporters and local groups across the UK.

Executive Summary

It might seem strange to some MPs that anyone could be opposed to 'Online Safety'. ORG supports sensible measures to protect children online. But the Online Safety Act (OSA), as currently written and interpreted, is already producing harmful unintended consequences for privacy, cybersecurity, free expression and the wider UK digital economy.

This briefing outlines concerns people have with the way the Act is working in practice and why they are upset with how age assurance has been introduced, both unsafely and applied to a wide range of content and with the wrong social media posts being censored. Small sites have been closed for fears about compliance with the Act. Such problems led to over 550,000 people signing a petition asking for the act to be repealed, which Parliament is considering on 15 December.

The Online Safety Act will always be limited in its ability to tackle online harms, because it focuses on removing illegal and harmful content rather than tackling the underlying economic and structural dominance of major platforms; it leaves the monopolistic business models, algorithmic prioritisation, and lack of interoperability which drive misinformation, polarisation, and loss of user control, largely untouched. These need to be tackled through strong market interventions, to place users in control of what content they receive and how.

We urge MPs to support a more balanced, evidence-based and rights-respecting approach that tackles the underlying causes of social media harms and protects children without harming all of our Article 10 rights to freedom of expression or Article 8 rights to privacy.

Main recommendations for Parliament

1. Regulate age assurance providers
2. Exempt small and low-risk services from the full weight of regulation.
3. Strengthen due-process protections to prevent wrongful automated takedowns and infringements of freedom of expression rights.
4. Require Ofcom's Codes to meet human rights and proportionality standards.
5. Protect VPN use and encryption, rejecting any enforcement strategy that undermines cybersecurity.
6. Use existing competition powers so that users can choose their content prioritisation and moderation engines, and switch their social media provider, without losing their networks of contacts, to drive better social media.

1. Why are people upset about content being age-gated?

There is evidence that most of the British public support pornographic content being age gated[1]. However, the Act does not just cover pornographic material nor does it clearly define the type of content that should be age gated.

Ofcom's Protection of Children Codes explicitly require platforms that rely on age-based denial of access, in order to remain safe, to know, or make a reasonable attempt to know, whether a user is a child through age assurance, which can be age verification or age estimation[2]. As such, platforms are using age-gating as a way to restrict under 18s' access to all sorts of content in order to avoid legal liability under the Act. It is also not clear or easy to know or categorise what sort of content should be placed behind an age-gate.

Because platforms face compliance costs, reputational damage and heavy penalties for non-compliance, they often apply age gating more broadly than strictly necessary. This includes content that is legally safe for children but carries any perceived risk. As a result, even borderline or lawful content may be placed behind an age gate, creating greater restrictions online content than for other forms of media, and turning age gating into a default safety measure, rather than a targeted means to prevent access to pornographic material.

People are therefore concerned that they are being asked to go through intrusive age-assurance processes on platforms like BlueSky, Spotify, Xbox gaming services, or to see certain Subreddits, that might be suitable for teenagers.

Without legal limits on how age assurance works, platforms and third-party vendors have an economic incentive to collect more data than necessary. Platforms also have an incentive to choose cheaper and less secure vendors, mainly located in the US, with poor data protection practices. Some of this data collection can cause more online harms to people through increased risk of cybercrimes.

For example, poorly implemented age assurance solutions have exposed users to new harms with the ID photos of up to 70,000 users on Discord leaked in a data-breach.[3] The ICO and current data protection regime is proving ineffective in regulating the industry. Additionally Ofcom are having to play 'whack-a-mole' enforcement with mirror sites that pop up without any age gating of content because the system relies on a platform implementing the law, rather than a

---

1How have Britons reacted to age verification? YouGov poll https://yougov.co.uk/technology/articles/52693-how-have-britons-reacted-to-age-verification

2Quick guide to children's access assessments - https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-access-assessments

3'ID photos of 70,000 users may have been leaked, Discord says' -BBC News October 2025 https://www.bbc.co.uk/news/articles/c8jmzd972leo

software or hardware control on a young person's device.

Platform versus device or operating system level age assurance

In recent weeks a debate has emerged about whether age assurance should occur at the level of the device rather than each platform. In practice it is very difficult for Ofcom to ensure every single platform complies with the law. Bad actors can quickly and easily create mirror sites of reputable platforms.

Open Rights Group believes that service providers should allow users to choose the method and identity provider they use. It may be possible for App stores on users' own devices to handle verification, as user profiles on personal devices already store extensive sensitive personal data locally under users' own control; however the risks of further centralising market control by Apple and Google would need to be addressed.

Could Digital ID solve this?

Using a Digital ID to prove your age would not solve the problem of the wrong sort of content being placed behind an age gate. To solve that problem Parliament needs to clearly and explicitly define the type content that should be behind an age 18 age-gate, and prohibit age gating for other forms of content.

Using a Government Digital ID to establish your age online could also come with other privacy risks if a Digital ID app is designed in a way in which it could 'phone home' and alert the Government when it had been used for such a purpose. It is unclear that users would accept the idea of using government ID to access adult content.

ORG recommends that:

- Parliament narrowly defines when age assurance is required.
- Platforms should provide users with detailed documents regarding the use of their data so that they can understand the risks to their privacy and data.
- Ofcom and the ICO should work with industry to create a high standard for privacy in age verification.
- Ofcom should recommend that age verification solutions include the use of high, independently managed data protection standards, and meet interoperability and accessibility needs.
- Future legislation should incorporate privacy, accessibility, and interoperability requirements for age verification and assurance.
- Users should have a choice of which age-assurance system they wish to use,

including a choice to use a device level proof of age.

For more detail on ORG's recommended statutory framework for age assurance, see our dedicated briefing.

2. Why people are upset their posts are being wrongly moderated or age-gated by the Act

There are some very clear reasons why the act is causing posts to be placed behind an age gate. This is not merely a case of platforms implementing duties created by the law badly. The causes for the problem lie in both the wording of the act and Ofcom's code of practices:

Strong financial penalties. The Act allows Ofcom to fine non-compliant services up to ten per cent of qualifying worldwide revenue or to block services in the UK for serious non-compliance.[4]

Broad risk reduction duties. For user to user services likely to be accessed by children, the Act requires a suitable children's risk assessment and ongoing measures to mitigate identified risks.[5]

Vague definitions of harmful content. The Act defers to Ofcom guidance and Codes for definitions of content harmful to children, giving services broad discretion over what is treated as harmful.[6]

Pressure to demonstrate proactive compliance. Platforms must implement design, operation, and mitigation measures including automated moderation, age assurance, gating, and access controls.

Ofcom codes recommending pre-emptive measures. The Protection of Children Code of Practice requires highly effective age assurance where high risk content is not prohibited for all users.[7]

The result is that, as a platform must know it is restricting "high risk content" for children, it is easiest to age gate parts of the service where that content might be found – such as through direct messages, or on forums or groups that discuss topics like alcohol, sex or drugs. This then restricts advice and helps that may be vital for young people to access.

4 Online Safety Act 2023, sch. 13, para. 4.
5 Online Safety Act 2023, Pt 3 Ch 2 ss 11–12.
6 Online Safety Act 2023, ss 60-61 (with Ofcom guidance per s. 53).
7 Ofcom, *Guidance to Proactive Technology Measures* (Draft, June 2025) Online Safety Act 2023, s. 231 (definition of "proactive technology") and Sch 4 para 13 (constraints on its use for analysing user-generated content).

3. Why people are upset their posts are being wrongly moderated

The same issues apply to content controls. While platforms have a duty to "have particular regard" to free expression,[8] this is a weak commitment compared to the drive to remove risks, in the face of fines.

There is a low threshold for content removal. Platforms only need to "reasonably suspect" that content is illegal before removing it.[9] Because the Act does not define illegal content in a way that automatically prescribes censorship, users cannot know in advance whether their content will be removed. This means removals are driven by platform discretion rather than clear legal rules, making it impossible to assess whether each removal is proportionate from a rights perspective, including freedom of expression.[10]

It is the OSA's structure that is therefore causing platforms to create moderation policies that are overly restrictive. This results in:

- legitimate speech being taken down;
- prior restraint censorship of content before it is published;
- political and journalistic content being suppressed;
- disproportionate harm to LGBTQ+ users, minority communities, and activists;
- little recourse or appeal for ordinary users;
- conflict with countries such as the US where speech is protected under their first amendment rights; and
- self-censorship as people develop a proxy language and deploy symbolism and slang to avoid AI moderation such as 'unalive, grape, slice vibes, yeeted'.

In addition, AI-based moderation is error-prone, biased, and lacks transparency. It is incredibly technically difficult to categorise and correctly moderate the scale of user content generated daily on social media platforms.[11] Some MPs have identified the problem with AI moderation already and an EDM has been tabled on the issue.[12]

---

8 Online Safety Act 2023, Pt 3 Ch 3 s 2

9 Online Safety Act 2023, ss 59, 193, 10. And for discussion of the "reasonable grounds" threshold and discretion risks: Ofcom, Illegal Content Judgements Guidance — threshold of "reasonable inference".

10 The Online Safety Act: proactive illegality duties, safeguards and proportionality, Graham Smith July 2025 https://www.cyberleagle.com/2024/07/the-online-safety-act-proactive.html

11 How to Fix the Online Safety Act: A rights based approach, Pt 2 Ch 5 - https://www.openrightsgroup.org/app/uploads/2025/05/How_to_fix_the_Online_Safety_Act_A_Rights_first_approach.pdf

12 Oversight of automated moderation by social media companies - EDM (Early Day Motion) 1858 - https://edm.parliament.uk/early-day-motion/64262

Unfortunately the harms associated with incorrect moderation of content will increase dramatically under the OSA, especially as Ofcom starts to require platforms deploy perceptual hash matching for a wider range of platforms, and user-to-user services and Ministers expand the number of priority offences via use of statutory instruments.

ORG recommends that Parliament:

- Strengthens the legal duty on platforms to consider freedom of expression from "have regard to" to "take all reasonable steps to protect".
- Introduces a proportionality test, for example: "A provider must take all reasonable steps to protect freedom of expression, consistent with the need to prevent illegal or harmful content."
- Provides clear routes for appeal, including to the courts, and correction for wrongful takedowns that considers the actual legality of content, not just whether it was 'reasonable to infer' the illegality of the content.
- Defines "online harms" with greater specificity to avoid ambiguity.
- Removes powers for the Secretary of State to unilaterally add priority offences to the Act.
- Requires Ofcom remove references to "bypass strategies" that encourage prior restraint censorship
- Strengthens standards for transparency, accuracy, and proportionality of moderation.

4. Local regulation, global consequences

Without treaties in place the UK has limitations to the extent which it can seek to fine companies in different jurisdictions. The Online Safety Act has already resulted in diplomatic tensions with the US, with the US House Judiciary Hearing on 'Europe's Threat to American Speech and Innovation'. This is because in the US all citizens have a first amendment right to free speech, and the act seeks to regulate US platforms.

Without reform to the law that respects the rights of citizens in other countries and their freedom to determine their own approaches to speech the act will continue to harm to our international relations.

5. Why are people now talking about VPNs

When age-assurance was introduced there was evidence that VPN apps surged in

downloads.[13] VPNs reduces exposure to cyberattacks and helps maintain consistent access to cloud services or region-specific tools. In addition to all these features a VPN can make a website think you are connecting from a different location by routing your traffic through another country. This can be used to circumvent UK specific age-gates.

The latest research evidence on VPN use is that the increase in use has come from adults rather than from children.[14] On 4 December, Ofcom's Online Safety Group Director told the BBC's Today Programme that UK VPN use had risen from 600,000 to well over one million people. Ofcom are now reporting that since August 2025 VPN use has been in decline and now sits around 900,000 users.

What is Ofcom doing about this?

As a result of the Online Safety Act, Ofcom has been paying the commercial data-broker Apptopia to obtain information on UK citizens' private software use.[15] This consumer-level surveillance appears intended to assess national patterns of VPN adoption. Ofcom have been under pressure to monitor VPN use from the Children's Commissioner and the House of Lords Communications and Digital Committee, both of whom have highlighted VPNs as a potential circumvention method.

Why trying to regulate professional VPN use is a bad idea

Regulating VPN use in the UK would be both impractical and harmful. A conservative estimate places the UK VPN market at £2.26 billion in 2025, with usage primarily by businesses and adults for cybersecurity, secure remote access, and managing networks with restricted sites, not as a widespread tool for children.[16]

Even the most authoritarian governments, such as China and Iran, struggle to control VPNs, demonstrating the futility of such regulation. Forcing websites to try and detect VPN use would cause widespread disruption to users in other countries.

Data protection regulation that restricted the commercialisation of people's browsing data through VPNs however could be an effective way to stop free VPNs exploiting people's personal data. Free VPNs are also far more accessible to children as there is no cost or credit card requirement to obtain one.

---

13 UK VPN demand soars after debut of Online Safety Act https://www.theregister.com/2025/07/28/uk_vpn_demand_soars/

14 New research from Childnet shows that the 'surge' in VPN use following the introduction of age verification in the summer is not attributable to children https://saferinternet.org.uk/blog/new-research-from-childnet-into-vpns

15 Exclusive: Ofcom is monitoring VPNs following Online Safety Act. Here's how https://www.techradar.com/vpn/vpn-privacy-security/exclusive-ofcom-is-monitoring-vpns-following-online-safety-act-heres-how

16 Spherical insights, https://www.sphericalinsights.com/reports/united-kingdom-virtual-private-network-market

[If you want to learn more about VPNS and the Online Safety Act we have produced a specific briefing on this issue]

6. Why people fear the Online Safety Act could dismantle meaningful end-to-end encryption (E2EE)

End-to-end encryption (E2EE) is a fundamental cybersecurity measure: it ensures that only the sender and the intended recipient can read messages or view shared media. It does this by encrypting a message on a device and then only decrypting it on the receiving device. It prevents interception of private chats on services like WhatsApp, Facebook Messenger or Signal. For many people, including parents sharing photos privately with family, E2EE provides essential protection for privacy and security, guarding against risks such as identity theft, targeting, or misuse of personal media. The Information Commissioner's Office (ICO) has publicly defended E2EE, arguing that it "serves an important role both in safeguarding our privacy and online safety," including by preventing criminals from accessing children's pictures or location details.[17]

Under the Online Safety Act, however, the regulator Ofcom has identified "encryption" as a potential "risk factor" for online harms. The concern: E2EE could enable predators to exchange harmful content. including illicit images or grooming messages. The worry is this could occur out of sight of both platforms and law enforcement. As a result, the law grants Ofcom the power (via a "technology notice") to require services to seek to develop new unproven technologies such as client-side scanning[18] to scan private messages, even where encrypted, for illicit content like child sexual abuse material (CSAM) or terrorist content.

In practice, implementing such scanning on E2EE platforms would likely require dismantling or weakening encryption (for example via back-doors or client-side scanning), undermining the very privacy and security E2EE was designed to guarantee. That would expose all users, not just those allegedly abusing the system — to increased risk: from hackers, data breaches, mass surveillance, or economic harms (e.g., compromised corporate communications, banking, and trade secrets).[19]

In the UK, the Police already have other means of accessing evidence: if a device is seized during an investigation, messages stored on it decrypted on the device by

17 End-to-end encryption protects children, says UK information watchdog https://www.theguardian.com/technology/2022/jan/21/end-to-end-encryption-protects-children-says-uk-information-watchdog

18 Bugs in our pockets: the risks of client-side scanning, *Journal of Cybersecurity*, Volume 10, Issue 1, 2024, https://doi.org/10.1093/cybsec/tyad020

19 Breaking encryption is legally and practically unworkable https://www.indexoncensorship.org/2025/11/new-report-breaking-encryption-is-legally-and-practically-unworkable/

design can be read. If someone refuses to allow access to an encrypted device they can be charged under Section 49 of RIPA, and if their device is being searched at the border they can be charged under Section 7 of the Terrorism Act. Weakening E2EE is not the only route to obtaining evidence.

Weakening encryption in a broad, systemic way therefore jeopardises cybersecurity and privacy for millions, in pursuit of potential but uncertain gains in detecting criminal content. An approach that undermines cybersecurity cannot be considered a credible "online safety" regime.

ORG recommends the following changes to the Act:

- Remove "technology notices" entirely.
- Prevent such notices from applying to private messaging services.
- Amend the Act to require judicial (not just "skilled person") oversight for any attempt to compel scanning or decryption.
- Require an assessment of the economic and cybersecurity harms that could arise from requiring scanning of E2EE messages or files.
- Remove encryption from Ofcom's "risk register".

7. Why people who run small low risk community sites and organisations like Wikipedia are upset with the act

The OSA applies burdens designed for global platforms to:

- small businesses
- community forums
- charities and clubs
- blogs and hobby sites
- open-source projects
- federated platforms (eg, Mastodon instances)
- public-interest services like Wikipedia

Many cannot meet the Act's compliance demands, especially around risk assessments, moderation requirements and age assurance. The result is predictable: they geoblock UK users or shut down entirely. Wikimedia has warned it might have to introduce a cap on UK users if forced to comply with regulations designed for large social media platforms.[20]

20 Wikipedia threatens to limit UK access to website, Telegraph, July 2025 -

ORG is tracking this emerging trend on our Blocked website.[21]

This reduces diversity online, harms SMEs, and accelerates market consolidation into the largest platforms.

ORG has drafted a proposed amendment to the Online Safety Act that would exempt small low-risk sites while retaining Ofcom's ability to enforce small high risk sites, by introducing an exemption for very small, not-for profit services like community forums, that are objectively low risk (in the view of a "reasonable person") while also allowing Ofcom to compel any such site to comply with the Act's usual duties, should Ofcom believe the site is in fact risky. We are happy to speak with any parliamentarian who is interested this reform to the Act.

8. Should MPs do more to regulate AI Chat bots?

Recent high profile cases of AI Chat Bots harming children has placed AI chat bots under more scrutiny. Evidence on the impact of AI chat bots is still emerging and is mixed. Undoubtedly some is very concerning, especially regarding the encouraging or sycophantic tone AI chatbots use, which is misleading; and also, due to the inability of AI tools to handle context, the default presumptions built into AI tools that discussion of sex, racism and other difficult topics is inherently harmful, leading to misleading summaries and searches.

If regulation is extended to AI chatbots, then the Online Safety Act would need to be updated. If it were to do so then MPs should at the same time address some of the privacy and freedom of expression concerns with the act.

This would avoid repeating the same problems we are experiencing with the Online Safety Act currently being applied to AI Chat Bot services.

Risk to poorly regulating Chat Bots

- There may be demands for stricter age-verification or usage restrictions for under-18s. That could mean children would need to prove age (ID, facial scan, etc.) to use chatbots. This could deny teenagers access to any social benefits that could arise from this technology.
- Content filters are already aggressive, lead to misleading results and could become more aggressive. Rather than just removing illegal content, chatbots may be required to proactively block or refuse content that might be "harmful to children" leading to further conservative moderation and over blocking.

---

https://www.telegraph.co.uk/business/2025/07/23/wikipedia-threatens-limit-access-website-britain/

21 https://www.blocked.org.uk/osa-blocks

- The development of new chatbots or alternative, smaller services could be hampered due to compliance costs and legal risk. This would reduce innovation and diversity in the AI space, benefiting big players through a form of regulatory capture.

9. A better, evidence-based approach exists

ORG is not outrightly opposing the Act. We support safer online environments, but through measures we think would work more effectively including:

- **Platform accountability**, including transparency and audits
- **User empowerment**, including filtering tools and better reporting
- **Proportionate rules** focused on the highest-risk services
- **Privacy-preserving design**
- **Protection for encryption and cybersecurity**
- **Interoperability,** to ensure consumer competition can create safety.

Our report *Making Platforms Accountable, Empowering Users and Creating Safety* highlights how government and society continue to fuel monopolies through advertising expenditure and policy dependence on major platforms. It urges the UK to apply its new Digital Markets, Competition and Consumers Act 2024, enabling the Competition and Markets Authority (CMA) and its Digital Markets Unit (DMU) to impose interoperability and data-portability obligations on firms with Strategic Market Status.

10. Further reading and reports from ORG

Making Platforms Accountable: Empowering users and creating safety

https://www.openrightsgroup.org/publications/making-platforms-accountable-empowering-users-and-creating-safety/Making Platforms Accountable: Empowering users and creating safety

Briefing VPNS and the Online Safety Act

https://www.openrightsgroup.org/publications/briefing-vpns-and-the-online-safety-act/

Regulating Age Assurance

https://www.openrightsgroup.org/publications/regulating-age-verification/

How to fix the Online Safety Act: A rights first approach

https://www.openrightsgroup.org/publications/how-to-fix-the-online-safety-act-a-rights-first-approach/

Credits