

Safety Last: The Online Safety Act and the Return of Unworkable Tech Mandates

Novalytics Gibraltar

Abstract—The Online Safety Act (OSA) represents a significant shift in the regulation of online content in the United Kingdom, introducing a statutory duty-of-care model with broad applicability to online services. While publicly associated with the regulation of pornographic material, its provisions extend to a wide range of non-pornographic platforms, creating substantial compliance, operational, and reputational implications. This paper examines the Act’s scope, the specific risks it poses to services outside the adult content sector, and the technical challenges inherent in its enforcement. Drawing on historical precedents such as the United States’ Prohibition era, the 1990s Clipper Chip programme, and contemporary proposals to weaken end-to-end encryption, the analysis highlights recurring patterns of circumvention, security degradation, and erosion of public trust. The findings suggest that, absent technically robust and privacy-preserving enforcement mechanisms, the OSA risks replicating the failures of its historical analogues—undermining its own policy objectives while imposing significant burdens on compliant organisations.

Keywords—Cybersecurity, Policy, Data Protection, Online Safety Act, Encryption, Clipper Chip, End-to-End Encryption, Censorship, Circumvention, Prohibition

Contents

- 1 Introduction** **1**
- 2 Overview of the Act** **1**
 - 2.1 Service categories 1
 - 2.2 Core duties 2
 - 2.3 Categorisation and thresholds 2
 - 2.4 Enforcement powers 2
- 3 Implications for non-pornographic companies** **2**
 - 3.1 Age assurance spill-over 2
 - 3.2 Compliance overhead 2
 - 3.3 Reputational and operational impacts 2
 - 3.4 Chilling effects on innovation 2
- 4 Technical enforcement challenges** **2**
 - 4.1 The Prohibition analogy 2
 - 4.2 Comparison with the Clipper Chip 3
 - 4.3 Breaking end-to-end encryption 3
 - 4.4 History repeats 3
- 5 Recommended actions for affected companies** **3**
 - 5.1 Conduct a scope assessment 3
 - 5.2 Undertake a risk assessment 3
 - 5.3 Review and update policies 3
 - 5.4 Evaluate technical measures 4
 - 5.5 Establish compliance governance 4
 - 5.6 Monitor regulatory developments 4
- 6 Conclusion** **4**
- References** **4**

1. Introduction

The Online Safety Act 2023 (OSA) represents the most comprehensive attempt to regulate online content and user interactions to date. The legislation passed through multiple consultations and revisions before receiving Royal Assent on 26 October 2023 [1]. Its stated objectives are to protect children and adults from harmful online material, reduce the prevalence of illegal content, and impose clear legal duties on online platforms with links to the UK [2].

Unlike earlier Internet regulation initiatives in the UK, the OSA adopts a duty-of-care model. In order to change this mindset, the accountability has now been shifted to online service providers who must actively assess risks, implement mitigation measures, and document compliance [3]. This obligation applies not only to platforms that host user-generated content, but also to search services, with additional provisions for pornography providers under Part 5 of the Act [4].

Although public discourse has focused heavily on the impact on adult content providers, the scope of the OSA is significantly broader. The Act applies to any service with a “link to the UK”, including platforms that are not based domestically, if they have UK users or target UK audiences [5]. As such, companies without any involvement in pornographic material may still be required to introduce age assurance measures, implement content moderation workflows, and comply with transparency reporting obligations [6].

The implementation of the Act is phased, with the first set of duties on illegal content already in force and additional requirements for child safety, pornography access controls, and categorised service obligations scheduled during the 2024 - 2026 period [7]. Ofcom, as designated regulator, has published codes of practice, enforcement guidelines, and compliance timetables to support service providers in meeting their obligations [3]. However, the breadth of scope, technical demands and potential conflicts with privacy-preserving technologies have raised significant concerns in the technology sector, academia, and civil society [8], [9].

The OSA is part of a wider global trend towards imposing statutory duties on online intermediaries. In the European Union, the Digital Services Act (DSA) establishes similar risk assessment, content moderation, and transparency obligations for platforms, with stricter measures for very large online platforms and search engines [10]. The Australia Online Safety Act 2021 grants its eSafety Commissioner powers to require removal of harmful material and enforce age restriction measures [11]. In Canada, the proposed Online Harms Act would create a Digital Safety Commission to regulate online content [12], while in the United States, legislative proposals such as the Kids Online Safety Act (KOSA) seek to impose child protection and age verification requirements [13].

Although each jurisdiction varies in scope and enforcement, these measures share common characteristics: an emphasis on proactive monitoring, a focus on protecting minors, and significant penalties for noncompliance. However, the OSA is notable for its combination of broad applicability, prescriptive enforcement powers, and specific targeting of pornography access controls [2].

2. Overview of the Act

The OSA establishes a statutory framework for regulating on-line services that facilitate user interaction or host user-generated content. It applies to three main service categories: user-to-user services, search services, and Part 5 providers of pornographic content [5]. The territorial scope of the Act is broad, covering services with any “link to the UK”, including those based overseas [1].

2.1. Service categories

- **User-to-user services:** Platforms that allow users to encounter content generated, uploaded, or shared by other users, such as social media networks, messaging services, and online marketplaces with user review functions [3].

- **Search services:** Providers that index, rank and return Internet results, including general-purpose search engines and on-platform search functions [5].
- **Part 5 providers:** Services that publish or make available pornographic material to UK users, regardless of whether this is their main function [4].

2.2. Core duties

The Act imposes a series of duties of care designed to reduce the prevalence of illegal and harmful content. These include [2], [3]:

- **Illegal content duties:** Taking proactive measures to prevent the presence and rapid dissemination of prior illegal content, such as terrorism, material for child sexual abuse and certain fraudulent activities.
- **Child safety duties:** Applying risk assessment and mitigation measures to protect children from harmful but not necessarily illegal content.
- **Adult safety duties:** For Category 1 services (as designated by Ofcom), providing adult users with tools to manage exposure to harmful content.
- **Risk assessment and transparency:** Conduct regular risk assessments, document mitigation strategies, and publish transparency reports on content moderation and enforcement actions.
- **Record-keeping and governance:** Maintain compliance records, appointing responsible individuals, and establishing clear internal processes to respond to Ofcom information requests.

2.3. Categorisation and thresholds

Ofcom will designate certain services into categories based on metrics such as user numbers, functionality, and risk profile [14]. Category 1 services—typically very large platforms—will face the most stringent requirements, particularly around adult safety and user empowerment tools. Category 2A and 2B services will be subject to reduced duties, but must still comply with illegal content and child safety obligations.

2.4. Enforcement powers

Ofcom has significant enforcement powers under the OSA, including the ability to [15]:

- Fines of up to £18 million or 10% of annual global turnover, whichever is higher.
- Apply service restriction orders through the courts, compelling internet service providers and app stores to block access to non-compliant services.
- Require the use of specific technologies, such as accredited age-assurance systems, to meet statutory duties.

The combination of a broad territorial reach, detailed statutory duties, and strong enforcement powers distinguishes the OSA from many comparable regulatory regimes. Although this design aims to ensure robust compliance, it also creates significant compliance burdens for organisations, including those whose services do not directly relate to high-risk content categories.

3. Implications for non-pornographic companies

Although the OSA’s public profile has been dominated by its regulation of pornographic content, the legislation has far-reaching implications for organisations operating outside of the adult content sector. This is a direct result of the broad definitions of services in the Act, its “link to the UK” territorial scope, and its duty-of-care framework [2], [5].

3.1. Age assurance spill-over

A key area of concern is the potential spillover of age-assurance obligations to services without a focus on sexual or explicit content. The

statutory test for whether a service is “likely to be accessed by children” does not depend on the intended audience of the service, but on its actual or reasonably foreseeable user base [16]. As a result, online games, discussion forums, educational platforms, and community applications may be required to implement age-verification or age-estimation technologies, even where their content is predominantly benign.

Such measures could necessitate:

- Integration with third-party age assurance providers.
- Collect of additional personal data for verification purposes.
- Increased risk of user friction, leading to potential loss of engagement.

3.2. Compliance overhead

Nonpornographic services that fall within the scope of the Act may be required to undertake the same compliance activities as high-risk content providers. These include:

- Conducting and documenting regular risk assessments [6].
- Maintain detailed records of content moderation activities.
- Produce periodic transparency reports for public and regulatory scrutiny.
- Appointing named individuals responsible for compliance with potential personal liability in the event of breaches.

For smaller companies and start-ups, these obligations could represent a substantial operational and financial burden, particularly where compliance requires the deployment of new moderation technologies or expanded trust-and-safety teams.

3.3. Reputational and operational impacts

The introduction of intrusive verification measures and expanded content moderation may have unintended reputational effects. Privacy-conscious users may object to age-assurance processes, particularly those that require government-issued identification. Furthermore, content moderation errors—whether over-removal or under-removal—could result in public criticism or legal disputes. The possibility of Ofcom enforcement action, including blocking services, introduces an additional operational risk to businesses relying on uninterrupted access to the UK market [15].

3.4. Chilling effects on innovation

Mandatory risk assessments, moderation tooling, and age-verification systems can discourage smaller providers from launching in the UK market. This phenomenon, often termed a *chilling effect*, has been observed in other jurisdictions with strict intermediary liability or content control laws [8], [9]. The resulting reduction in market diversity could disadvantage UK users, limiting competition and choice in online services.

4. Technical enforcement challenges

From a technical point of view, several provisions of the OSA face fundamental implementation challenges that echo the failures of earlier regulatory and technological interventions in digital communications.

4.1. The Prohibition analogy

Critics have compared OSA enforcement ambitions with the prohibition era of the United States (1920–1933), in which a legal ban on alcohol production, distribution, and consumption inadvertently fuelled the growth of organised crime, unregulated production, and illicit distribution networks [17]. The analogy reflects a core enforcement paradox: When legal restrictions fail to align with the practical realities of user behaviour and technology, compliance is often bypassed through circumvention. In the context of OSA, determined

users can employ virtual private networks (VPNs), decentralised hosting, TOR browsing solutions, and peer-to-peer content sharing to evade platform-level restrictions [18].

Historically, such programmes have consistently failed in practice. The following sections explore the similarities between OSA, the Clipper Chip initiative, and proposals to weaken end-to-end encryption.

4.2. Comparison with the Clipper Chip

The Clipper Chip programme, announced by the United States government in 1993, was an attempt to reconcile the growing use of strong encryption with the desire of law enforcement for lawful access to communications. The initiative proposed embedding a hardware-based encryption system in communications devices, incorporating a “law enforcement access field” (LEAF), a key escrow mechanism that enables authorised government agencies to decrypt communications when presented with the necessary legal authority [19]. Under this scheme, the encryption keys for each device would be split and held in escrow by two separate government agencies, ostensibly to provide a safeguard against misuse.

Although the concept was presented as a balanced compromise between privacy and public safety, the programme encountered immediate and sustained criticism on multiple fronts.

- **Single point of failure:** The escrowed keys introduced a centralised vulnerability, susceptible to both insider abuse and external compromise. A breach of the escrow repositories would expose all communications encrypted by Clipper-enabled devices.
- **Technical incompatibility:** The mandate failed to account for the rapid proliferation of alternative, non-US encryption technologies. These could not be feasibly restricted to escrow-compliant systems, rendering the policy ineffective on a global communications network.
- **Security weaknesses:** Cryptographers quickly discovered flaws in the LEAF implementation, demonstrating that the system could be circumvented without triggering legal access, undermining its stated purpose.
- **Public and industry backlash:** Privacy advocates, technology companies, and civil liberties groups characterised the proposal as an unacceptable intrusion into personal privacy and commercial confidentiality, warning that it would set a precedent for government backdoors in other forms of digital security.

The Clipper Chip ultimately failed to gain market adoption and was formally abandoned by the late 1990s. However, its legacy endures as a case study in the limitations of technically intrusive mandates. It illustrates that backdoor-based access mechanisms, even when framed as lawful and controlled, can falter due to security vulnerabilities, incompatibility with global technological realities, economic impracticality, and overwhelming public opposition. The parallels with the potential impact of the Online Safety Act on end-to-end encryption are striking, highlighting the persistent tension between government access requirements and the fundamental principles of secure communications.

4.3. Breaking end-to-end encryption

Particularly contentious is the potential impact of OSA on end-to-end encrypted (E2EE) messaging services. Although the Act does not explicitly ban E2EE, it empowers Ofcom to require the use of “accredited technology” to detect illegal content, which could include client-side scanning or other pre-encryption inspection techniques [20]. This approach has been tried many times before and always results in technical measures being deployed to make message interception even harder for people that do not want their messages intercepted, but a lot easier for Governments to snoop into normal users traffic.

- **Security degradation:** Introducing scanning mechanisms on the client side or in the encryption pipeline creates additional attack surfaces.

- **Loss of confidentiality:** True E2EE ensures that only the sender and recipient possess the keys to decrypt a message; mandated inspection undermines this guarantee.
- **Global interoperability issues:** Messaging platforms operate in multiple jurisdictions, and complying with a UK-specific inspection mandate could require creating jurisdiction-specific forks (with inconsistent security properties) or weakening encryption globally.

4.4. History repeats

We can see the OSA is having the exact same problems as the proposals attempted before to monitor and police content, or to gain better intelligence from mass data extraction of internet data collection.

- **Circumvention inevitability:** As with Prohibition and the Clipper Chip, determined actors can and will migrate to unregulated, decentralised, or foreign-hosted services to bypass controls.
- **Security trade-offs:** Backdoor or inspection mandates introduce systemic vulnerabilities that can be exploited by malicious actors.
- **Erosion of trust:** Users and organisations may lose confidence in domestic services subject to intrusive scanning, incentivising the adoption of foreign or underground alternatives.
- **Jurisdictional limits:** Technical mandates tied to national law face severe enforcement limits in the context of a globally interconnected network.

In summary, while OSA goals are framed in terms of prevention of harm and user safety, the available technical methods for enforcement risk introducing vulnerabilities, undermining privacy, and replicating the failure patterns of previous interventions.

5. Recommended actions for affected companies

In light of the broad applicability of the OSA and the technical enforcement challenges identified, organisations providing accessible online services in the UK should take proactive steps to address both compliance obligations and operational risks. Although we consider these requirements to be completely a misguided approach by the UK Government, the reality is that all businesses must still ensure compliance with the law. The following measures outline practical steps organisations should consider implementing to safeguard their operations and minimise regulatory exposure.

5.1. Conduct a scope assessment

Determine whether the service is a regulated user-to-user or search service and whether it has a “link to the UK” as defined in the Act. Assess whether the service is “likely to be accessed by children” and whether any characteristics could trigger additional duties. Document the assessment and review it as products, audiences and Ofcom guidance evolve [1], [2], [5].

5.2. Undertake a risk assessment

If in scope, perform an assessment of child safety risk and illegal content (where applicable) aligned with Ofcom’s templates and Codes of Practice. Identify relevant content risks, high-risk features (for example, live streaming, open DM) and the proportional mitigations that you will apply [3], [6].

5.3. Review and update policies

Update acceptable use, reporting, appeals, and privacy notices to reflect OSA duties and user rights. Where age assurance is introduced, address data minimisation, retention, and user transparency; ensure consistency with Ofcom’s child safety materials and your published safety features [2], [16].

5.4. Evaluate technical measures

Assess the feasibility and user impact of moderation workflows, reporting tools, and detection technologies referenced in Ofcom's codes. Where encryption is used, analyse the implications of any detection or audit directions and record a security risk trade-off; avoid weakening end-to-end protections without a proportionate, regulatory-backed approach [3], [8], [20].

5.5. Establish compliance governance

Assign a named owner for OSA compliance, define escalation paths, and prepare Forcom information requests and audits. Map deliverables (e.g., transparency reporting cadence) to Ofcom's enforcement guidance and published compliance dates [7], [15].

5.6. Monitor regulatory developments

Track Ofcom consultations, category designations, timelines, and updated codes. Build a review calendar keyed to Ofcom's 'important dates' so that product and policy teams can adjust ahead of enforcement milestones [2], [7].

6. Conclusion

The Online Safety Act represents an ambitious attempt to address the complex problem of harmful and illegal online content through a statutory duty-of-care framework. Its scope extends well beyond high-risk sectors such as pornography, imposing substantial compliance obligations on a wide range of online services, including those with no direct connection to explicit material. The combination of broad territorial applicability, detailed risk assessment requirements, and potentially intrusive enforcement mechanisms creates significant operational, financial, and reputational challenges for affected organisations.

From a technical perspective, the Act faces the same limitations that have historically undermined other regulatory attempts to impose control over distributed, globalised communication systems. Prohibition illustrates the risks of legislating without aligning with practical behavioural and technological realities: where compliance is onerous or invasive, circumvention becomes a rational choice for users. The clipper chip demonstrated that introducing systemic access mechanisms or backdoors not only fails to achieve universal compliance, but actively degrades security and trust. Efforts to mandate scanning of end-to-end encrypted communications repeat these mistakes, introducing vulnerabilities and creating incentives for migration to unregulated or foreign-hosted platforms.

We have already seen that the top performing apps in the UK's Google App store are all VPN applications. Users are already circumnavigating the controls, and on a large scale [21]

Although OSA's policy objectives are framed around user safety and harm prevention, its technical enforcement strategies risk producing counterproductive outcomes: weakening security, undermining privacy, and erode trust in regulated services. In the absence of workable, privacy-preserving enforcement methods that can operate at scale and across jurisdictions, the Act may ultimately face the same fate as its historical attempts to monitor the unmonitorable: high-profile, politically costly, and largely ineffective in achieving its intended goals.

Contact Novalytics for More Information

Novalytics specialises in cybersecurity, information governance, and advanced analytics solutions designed specifically for SMEs operating within high-risk sectors. Our experts provide personalised guidance, strategic insight, and practical support to protect your organisation against evolving cyber threats.

For additional details on cybersecurity best practices, assistance with regulatory compliance, or a consultation on improving your organisation's cyber resilience, please contact us via:

- Website: <https://www.novalytics.co.uk>
- Email: contact@novalytics.co.uk

References

- [1] UK Parliament, *Online safety act 2023*, <https://bills.parliament.uk/bills/3137>, Accessed: 2025-08-11, 2023.
- [2] Ofcom, *Ofcom's approach to implementing the online safety act*, <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation>, Accessed: 2025-08-11, 2025.
- [3] Ofcom, *Codes of practice: Online safety act*, <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/codes-of-practice>, Accessed: 2025-08-11, 2024.
- [4] UK Government, *Online safety act 2023, part 5: Pornography*, <https://www.legislation.gov.uk/ukpga/2023/50/part/5>, Accessed: 2025-08-11, 2023.
- [5] UK Government, *Online safety act 2023, scope of application*, <https://www.legislation.gov.uk/ukpga/2023/50/section/4>, Accessed: 2025-08-11, 2023.
- [6] Ofcom, *Quick guide to illegal content risk assessments*, <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-online-safety-risk-assessments>, Accessed: 2025-08-11, 2025.
- [7] Ofcom, *Important dates for online safety compliance*, <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/important-dates-for-online-safety-compliance>, Accessed: 2025-08-11, 2025.
- [8] Electronic Frontier Foundation, *The uk online safety bill: A massive threat to online privacy, security, and speech*, <https://www.eff.org/pages/uk-online-safety-bill-massive-threat-online-privacy-security-and-speech>, Accessed: 2025-08-11, 2023.
- [9] Open Rights Group, *Online safety act: A guide for organisations working with the act*, <https://www.openrightsgroup.org/publications/online-safety-act-a-guide-for-organisations-working-with-the-act/>, Accessed: 2025-08-11, 2023.
- [10] European Union, *Regulation (eu) 2022/2065 on a single market for digital services (digital services act)*, <https://eur-lex.europa.eu/eli/reg/2022/2065>, Accessed: 2025-08-11, 2022.
- [11] Australian Government, *Online safety act 2021*, <https://www.legislation.gov.au/Details/C2021A00076>, Accessed: 2025-08-11, 2021.
- [12] Government of Canada, *Online harms act*, <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-63/first-reading>, Accessed: 2025-08-11, 2024.
- [13] US Congress, *Kids online safety act*, <https://www.congress.gov/bill/118th-congress/senate-bill/1409>, Accessed: 2025-08-11, 2023.
- [14] Ofcom, *Regulated service categories under the online safety act*, <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/overview-of-regulated-services.pdf?v=387540>, Accessed: 2025-08-11, 2024.
- [15] Ofcom, *Enforcement guidance for the online safety act*, <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/online-safety-enforcement-guidance.pdf?v=391925>, Accessed: 2025-08-11, 2025.

- 422 [16] Ofcom, *Quick guide to protection of children codes*, [https://](https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-safety-codes)
423 [www.ofcom.org.uk/online-safety/illegal-and-harmful-](https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-safety-codes)
424 [content/quick-guide-to-childrens-safety-codes](https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-safety-codes), Accessed:
425 2025-08-11, 2024.
- 426 [17] D. Oppenheimer, *The Economics of Prohibition*. University of
427 Chicago Press, 1995.
- 428 [18] Electronic Frontier Foundation, *Americans, be warned: Lessons*
429 *from reddit's chaotic uk age verification rollout*, [https://www.](https://www.eff.org/deeplinks/2025/08/americans-be-warned-lessons-reddits-chaotic-uk-age-verification-rollout?language=en)
430 [eff.org/deeplinks/2025/08/americans-be-warned-lessons-](https://www.eff.org/deeplinks/2025/08/americans-be-warned-lessons-reddits-chaotic-uk-age-verification-rollout?language=en)
431 [reddits-chaotic-uk-age-verification-rollout?language=en](https://www.eff.org/deeplinks/2025/08/americans-be-warned-lessons-reddits-chaotic-uk-age-verification-rollout?language=en),
432 Accessed: 2025-08-11, 2025.
- 433 [19] A. M. Froomkin, "Metaphor is the key: Cryptography, the clip-
434 per chip, and the constitution," *U. Pa. L. Rev.*, vol. 143, p. 709,
435 1994.
- 436 [20] UK Government, *Online safety act 2023, audit and detection*
437 *duties*, [https://www.legislation.gov.uk/ukpga/2023/50/](https://www.legislation.gov.uk/ukpga/2023/50/section/111)
438 [section/111](https://www.legislation.gov.uk/ukpga/2023/50/section/111), Accessed: 2025-08-11, 2023.
- 439 [21] L. McMahon, "VPNs top App Store charts as UK age verifica-
440 tion kicks in," *BBC News*, Jul. 2025. [Online]. Available: [https:](https://www.bbc.com/news/articles/cn72ydj70g5o)
441 [//www.bbc.com/news/articles/cn72ydj70g5o](https://www.bbc.com/news/articles/cn72ydj70g5o).