The Online Safety Act 2023: A Comprehensive Analysis of its Impact and Implications

1. Introduction to the Online Safety Act 2023

1.1 Legislative Context and Primary Objectives

The Online Safety Act 2023 (c. 50) represents a pivotal legislative achievement by the Parliament of the United Kingdom, having received Royal Assent on October 26, 2023.[1] Its core objective is to establish a robust and comprehensive regulatory framework for online content and services operating within the UK. The UK government has consistently championed this Act as a transformative development, asserting its aim to position the "UK the safest place in the world to be online".[3] This ambitious declaration underscores a perceived urgent need for stringent digital regulation.

The impetus for the Act originated from extensive government consultations on online harms, which revealed concerning statistics: a significant 61% of adults and a striking 79% of 12-15-year-olds reported experiencing at least one potentially harmful online incident within the preceding year.[3] These findings served as the primary justification for the legislative intervention, framing the Act as a necessary response to a pervasive societal challenge. The widespread public and governmental concern over these online harms directly fueled the political will and public acceptance for such a comprehensive and stringent piece of legislation. This framing also strategically legitimizes the Act's broad powers, making it politically challenging to argue against its core objectives despite potential drawbacks.

Beyond mere content removal, the strategic priorities underpinning the Act encompass a more holistic approach to digital well-being. These priorities include embedding "safety by design"

into online services, enhancing the transparency and accountability of platforms, maintaining regulatory agility to adapt to rapid technological advancements, fostering an inclusive and resilient online society, and promoting innovation in online safety technologies.[4] This indicates a recognition that simply deleting harmful posts is insufficient; the Act aims to address the systemic issues and underlying mechanisms that facilitate harm. This approach implies that compliance will require platforms to integrate safety considerations throughout their entire product lifecycle, from initial design and development to algorithmic deployment and user interface. It pushes for a more embedded and architectural approach to safety, moving beyond reactive moderation to proactive harm prevention at the foundational level of online service provision.

A core focus of the Act is the proactive identification and removal of illegal online content, including material related to terrorism, foreign interference, fraud, illegal abuse, threats, and hate crimes.[4] Paramount importance is placed on ensuring the strongest protections for children, with a relentless pursuit to reduce their exposure to harmful content. The Act also specifically targets illegal and misogynistic content to enhance online safety for women and girls.[4] Fundamentally, the Act signifies a profound shift in the UK's regulatory approach, moving from a reactive "notice and takedown" model, which primarily responded to reported harms, to a proactive "duty of care" framework, obligating online intermediaries to anticipate and prevent harm.[5] This is not a minor procedural tweak but a fundamental redefinition of regulatory responsibility. It implies that platforms are no longer just custodians of content but are now legally accountable for the

potential for harm arising from their services. This redefinition necessitates a significant re-engineering of platform operations, requiring substantial investment in predictive analytics, AI-driven content identification, and human moderation at scale. It transforms online services into highly regulated entities, akin to financial institutions, where risk management and preventative measures are paramount. This could set a global precedent for how governments expect platforms to operate, potentially influencing regulatory approaches in other jurisdictions.

1.2 Key Definitions and Scope of Application

The Act's expansive scope covers two primary categories of online services: "user-to-user services" and "search services".[2] "User-to-user services" are broadly defined as internet services where user-generated content—encompassing written material, messages, photos, videos, music, and data—can be encountered by other users.[2] This comprehensive definition includes, but is not limited to, social media platforms, content-sharing sites, messaging services, dating apps, online gaming platforms, and marketplaces. It even extends to ancillary features like comment sections,

online reviews, or discussion forums attached to core services.[3] "Search services" include platforms that enable individuals to search multiple websites, such as general internet search engines or specialized comparison sites.[3] The Act also explicitly includes "pornography providers" within its regulatory ambit.[6]

A critical aspect of the Act is its extraterritorial reach. The duty of care applies globally to services that have a "significant number of United Kingdom users," those that "target UK users," or those "capable of being used in the United Kingdom where there are reasonable grounds to believe that there is a material risk of significant harm".[2] This means foreign-based companies with a UK user base are subject to the Act. This broad and extraterritorial application is a deliberate design choice, aiming to prevent platforms from circumventing regulation by operating outside UK physical borders or by narrowly defining their services. This extraterritoriality creates a complex compliance challenge for global companies, potentially forcing them to either segment their services for the UK market or, more likely, apply the strictest UK standards universally across their global operations. This could lead to a "global floor" effect, where the most stringent national regulations effectively dictate global industry practice, impacting online content and services worldwide.

A key distinction from the EU's Digital Services Act (DSA) is the OSA's lack of explicit exemptions for small and micro enterprises.[7] Ofcom has unequivocally stated that "all in-scope user-to-user and search services, large or small, will need to take action to comply with the new duties".[3] This universal application stands in stark contrast to the tiered approach seen in the DSA. While this approach aims to ensure a universal baseline of online safety, it concurrently imposes a disproportionately heavy operational and financial burden on smaller entities, which often lack the extensive resources of larger tech companies. This situation risks stifling innovation among smaller players, potentially forcing them to withdraw from the UK market, as evidenced by some reported website closures [2], or leading to market consolidation where only large, well-resourced companies can afford the necessary compliance.

The Act primarily focuses on content and conduct that constitutes a *criminal offense* under UK law.[3] This differentiates it from the DSA, which encompasses a broader range of "illegal" and infringing content, including civil law infringements like defamation or intellectual property rights.[3] This distinction is a critical philosophical choice, narrowing the

legal scope of content regulation for adults compared to the DSA. This suggests a legislative intent to primarily address harms that are unequivocally illegal under existing UK criminal law, rather than venturing into more ambiguous areas of "harmful but legal" content for adults. While this might offer some clarity and reduce the risk of over-censorship of lawful adult speech, it also means that other forms of online harm (e.g., severe but non-criminal online bullying, pervasive misinformation that does not constitute a criminal offense) for adults may not be directly

regulated by the Act, relying instead on user empowerment features. The explicit broader category of "harmful" content for children highlights a specific legislative recognition of their unique vulnerability. While the Act aims to improve safety, Ofcom has clarified that it does not expect harmful and illegal content to be entirely eradicated, reflecting a risk-based approach to obligations.[3]

2. Core Provisions and Duties for Online Service Providers

2.1 The Duty of Care Framework

The cornerstone of the Online Safety Act is the establishment of a new "duty of care" for online platforms. This duty mandates that platforms take proactive action against illegal content and, crucially, against legal content that could be harmful to children where children are likely to access it.[2] This framework represents a significant departure from traditional regulatory models. Instead of merely responding to content once it has been reported (a reactive "notice and takedown" approach), the OSA requires platforms to actively monitor content. Where technically feasible, they must employ algorithms or other product features to prevent users from encountering clearly illegal content in the first instance.[5] This mandate for algorithmic accountability and systemic design responsibility extends legal responsibility beyond mere content moderation to the very

design and *functionality* of the online service itself. This is a direct push for "safety by design" [4], implying that platforms must integrate safety considerations into their core architecture, including how algorithms recommend, amplify, or curate content. It moves the regulatory focus from individual pieces of content to the systemic risks inherent in platform design, potentially leading to significant re-engineering of product development lifecycles and a greater emphasis on ethical AI development.

The duty of care applies broadly to all services within the Act's scope and encompasses several specific obligations: the illegal content risk assessment duty, the illegal content duties, the duty concerning rights to freedom of expression and privacy, the duties regarding reporting and redress mechanisms for users, and the record-keeping and review duties to ensure ongoing compliance.[2] While the Act explicitly includes "The duty about rights to freedom of expression and privacy" [2] as part of the duty of care, critics widely argue that the Act simultaneously poses a "significant threat to the right to privacy and freedom of speech and expression".[2] This highlights a fundamental and unresolved tension within the legislation. The Act attempts to legally mandate a balance between safety and rights, yet the practical implementation and interpretation of its

powers are proving highly contentious. Platforms face the complex task of fulfilling their duty to remove harmful content while simultaneously upholding freedom of expression and privacy, a balance that critics believe the Act's current framing makes exceedingly difficult, potentially leading to over-removal of lawful content.

2.2 Risk Assessment and Mitigation Requirements

A central pillar of the proactive duty of care is the requirement for service providers to conduct thorough illegal content risk assessments. These assessments must evaluate the likelihood of specific types of illegal content being available on their service and the risk that the service may be used to commit certain offenses.[3] Based on these assessments, companies are mandated to implement tailored mitigation measures. This transformation from content moderator to risk management entity demands that online platforms adopt sophisticated, enterprise-level risk management frameworks, traditionally seen in highly regulated sectors like finance or healthcare. This means platforms must invest heavily in expertise, data analytics, and operational processes to identify, assess, and continuously monitor potential harms. It shifts the operational focus from simply reacting to user reports to proactively identifying vulnerabilities in their systems and content flows, requiring a more integrated and continuous approach to safety.

These mitigation measures are wide-ranging and can include: establishing robust regulatory compliance and internal risk management arrangements; designing functionalities, algorithms, and other platform features with safety in mind; developing clear and enforceable terms of use policies; implementing policies governing user access to the service or particular content, including mechanisms for blocking users; employing effective content moderation practices, including content takedown procedures; providing functionalities that allow users to control the content they encounter; establishing comprehensive user support measures; and developing and implementing appropriate staff policies and practices.[3]

Ofcom is tasked with publishing detailed guidance and codes of practice to direct companies on these mitigation measures. Critically, compliance with these official codes will be treated as satisfying the statutory duties. However, companies that choose to implement alternative measures bear the burden of proving that their chosen methods are equally sufficient in meeting the Act's requirements.[3] While companies technically have the option to implement "alternative measures," the fact that compliance with Ofcom's published codes of practice automatically satisfies the duties creates a powerful incentive for platforms to adhere to these specific guidelines. Ofcom's guidance and codes will effectively become the industry standard for online safety practices in the UK. This centralizes regulatory influence and streamlines compliance

pathways for many companies, but it also means that Ofcom's interpretation of the Act will heavily shape how platforms design their services and manage content, potentially limiting innovative, but non-standard, safety solutions.

2.3 Specific Duties for Child Protection and Harmful Content

The Act places a strong emphasis on child protection, treating children as a "super-protected class" driving stringent measures. For services deemed "likely to be accessed by children," two additional and stringent duties are imposed: the children's risk assessment duties and the duties to protect children's online safety.[2] These duties require a wide-ranging assessment of risks arising from "content that is harmful to children," which includes "primary priority," "priority," and "non-designated harmful content".[3] This extensive and explicit additional set of duties, coupled with mandatory age verification, clearly demonstrates a legislative intent to treat children as a uniquely vulnerable demographic requiring the highest level of online protection. This creates a distinct, dual-tier regulatory system where child safety measures are significantly more stringent and less ambiguous than those for adults. It places immense pressure on platforms to accurately assess and verify user age, which has significant privacy implications and technical challenges. This focus also highlights the government's primary public justification for the Act's broad powers.

A particularly significant provision is Section 12, which mandates that service providers utilize age verification or age estimation technology to prevent children from accessing "primary priority content that is harmful to children." This category explicitly includes pornography, as well as content promoting, encouraging, or providing instructions for self-harm, eating disorders, or suicide.[2] This requirement applies to all services that make such content available, including social networks and internet pornography services, and self-declaration of age is deemed insufficient.[2] Beyond content filtering, the Act requires platforms to be designed with children's well-being in mind, offering clear reporting tools, easy-to-understand terms of service, and accessible support mechanisms when problems arise.[8]

For "category 1" services, designated as the largest global platforms (to be defined in secondary legislation based on risk and UK user base size [7]), four further duties apply: the adults' risk assessment duties, the duties to protect adults' online safety, the duties to protect content of democratic importance, and the duties to protect journalistic content.[2] Furthermore, Category 1 services have an additional duty to empower adult users by including features that allow them to increase control over the types of content they encounter, such as abusive content or content encouraging suicide/self-injury.[3] This emphasis on adult user empowerment as a complementary

regulatory tool for Category 1 services signifies a recognition that for adult users, the regulatory solution is not always outright content removal. Instead, it involves providing tools for users to curate their own online experience. This indicates a nuanced approach where the Act aims to protect children through proactive removal and age-gating, while for adults, it shifts towards transparency and user-driven customization of content exposure. This could lead to a proliferation of sophisticated user settings and content filtering tools becoming standard features across major platforms, potentially increasing user agency but also placing a burden on users to actively manage their safety.

2.4 New Communications Offenses

The Online Safety Act introduces several new criminal offenses, significantly expanding the scope of criminal law into the digital realm. This direct criminalization of specific online harms aims to deter malicious conduct. The creation of highly specific new criminal offenses demonstrates a legislative intent to move beyond platform liability and directly hold individuals accountable for particularly egregious and identifiable online harms. The identified prevalence and severity of these specific online harms directly led to their explicit criminalization, providing law enforcement with clear legal tools to prosecute individuals engaging in these malicious behaviors. This signals a proactive attempt to close legal loopholes that previously allowed such conduct to go unpunished.

The new offenses include:

- **"Epilepsy Trolling":** Section 183 creates an offense for "sending or showing flashing images electronically" if it is "reasonably foreseeable that an individual with epilepsy would be among the individuals who would view it," the sender intends to cause that person harm, and they have "no reasonable excuse." This provision is specifically designed to prevent malicious acts targeting individuals with epilepsy.[2]
- **Encouraging or Assisting Serious Self-Harm:** Section 184 makes "encouraging or assisting serious self-harm" a criminal offense, mirroring the existing offense of encouraging or assisting suicide. The first conviction under this section occurred in July 2025.[2]
- **False Communications:** The Act introduces an offense for sending false communications with the intent to cause non-trivial psychological or physical harm.[3]
- **Threatening Communications:** It creates an offense for sending messages that convey a threat of death or serious harm with the intent to cause fear.[3]

Additionally, the Act adds two new offenses to the Sexual Offences Act 2003:

- **Cyberflashing:** Sending images of a person's genitals.[2] The first conviction for

cyberflashing under this new law occurred in March 2024.[2]

- **Sharing or Threatening to Share Intimate Images:**.[2]

These new offenses effectively transpose and adapt traditional criminal law concepts (e.g., assault, incitement, harassment, sexual offenses) to the digital environment. They address behaviors that, while harmful, might have previously fallen into legal grey areas or been difficult to prosecute under existing statutes. This signifies a maturation of legal frameworks attempting to catch up with the evolving nature of online interactions and the harms they can facilitate. It also places an implicit burden on online platforms to identify and report instances of such criminal content or conduct, thereby integrating them more deeply into the broader law enforcement and justice system.

3. Regulatory Framework, Enforcement, and Accountability

3.1 Ofcom's Role and Powers

Ofcom, the national communications regulator, has been designated as the primary body responsible for the enforcement and supervision of the Online Safety Act.[2] This grants Ofcom unprecedented authority over the digital landscape in the UK, establishing it as a "super-regulator" with significant leverage over the digital economy.

Ofcom is endowed with a comprehensive suite of wide-ranging enforcement powers, including:

- **Blocking Access:** The power to block access to specific user-to-user services or search engines from within the United Kingdom. This can be achieved through interventions by internet access providers and app stores, effectively cutting off access to non-compliant platforms.[2] Importantly, Ofcom must apply to a court to obtain these Access Restriction Orders.[2]
- **Service Restriction Orders:** The ability to impose requirements on ancillary services that facilitate the provision of regulated services. Examples include services that enable fund transfers (e.g., payment providers), search engines that display or promote content, or services that facilitate the display of advertising (e.g., ad servers or ad networks).[2] These orders also require court application.[2]
- **Information Gathering Powers:** Extensive authority to request documents, conduct audits of platform operations, and interview staff members to assess compliance.[3]
- **Monetary Penalties:** The power to levy substantial fines, amounting to up to £18 million or 10% of an organization's annual worldwide turnover, whichever figure is higher.[2] This maximum penalty significantly exceeds that of the EU's Digital Services Act (DSA), which

is capped at 6% of total worldwide annual turnover.[3]

Ofcom's powers extend far beyond traditional content regulation. The ability to block access to services and, more critically, to issue "service restriction orders" against *ancillary* services like payment providers and ad networks, grants Ofcom direct leverage over the entire financial and operational infrastructure that supports online services. The magnitude of fines further amplifies this power. This effectively transforms Ofcom into a potent digital regulator with the capacity to significantly disrupt the business models and operations of non-compliant platforms. Compliance becomes not merely a legal obligation but an existential business imperative. This also sets a precedent for regulators to reach beyond direct service providers to their entire value chain, potentially influencing regulatory approaches in other sectors.

Ofcom has publicly stated its intention to adopt a "risk-based and proportionate approach" to enforcement. This means efforts will primarily focus on larger organizations and "big name" platforms, given their wider reach and higher potential for harm.[6] However, it is crucial to note that all in-scope services, regardless of size, are legally obligated to comply with the Act.[6]

A significant aspect of the regulatory framework is the power granted to the Secretary of State under Section 44 of the Act. The Secretary of State can direct Ofcom to modify a draft code of practice for online safety if deemed necessary for reasons of public policy, national security, or public safety.[2] Ofcom is legally obligated to comply with such directions. While intended to ensure alignment with broader government priorities, this power introduces a direct political dimension to what is ostensibly an independent regulatory function. This raises significant concerns about potential political interference in regulatory processes, particularly regarding content moderation and freedom of expression, as critics already fear government overreach.[2] It creates a mechanism for executive influence over the interpretation and application of the Act's duties, potentially undermining the perceived independence of the regulator.

3.2 Phased Implementation and Compliance Deadlines

The full operationalization of the Online Safety Act is contingent upon Ofcom issuing a series of detailed codes of practice and guidance, which are being rolled out through a structured, phased approach.[3] This strategic phased rollout is designed to manage the complexity of the new regulatory landscape and facilitate industry adaptation.

Key deadlines for compliance activities include:

- **December 2024:** Ofcom published the Illegal Content Risk Assessment Guidance and the

Illegal Harms Codes of Practice. This initiated the requirement for organizations to complete their initial risk assessments and implement corresponding mitigation controls by **March 31, 2025**.[3]

- **January 2025:** The Code on Age Assurance and Child Access Assessments was published. This requires organizations to assess whether children can and are likely to access their platform, with a deadline for completion set for **April 16, 2025**.[6]
- **March 2025:** Ofcom issued specific guidance focused on the protection of Women and Girls, addressing harms such as misogyny, harassment, and intimate image abuse.[6]
- **April 2025:** The Children's Risk Assessment Guidance and Children's Risk Profiles were published. This initiated a three-month deadline, requiring organizations to complete these mandatory assessments by **July 24, 2025**, if their Child Access Assessment indicated that children are likely to access their service.[6]

Ofcom has already commenced engagement with larger online service providers, advising them of these new requirements and their respective deadlines.[6] This early engagement with major players aims to set a precedent and ensure initial high-impact compliance.

4. Impact and Implications of the Online Safety Act 2023

4.1 Impact on Freedom of Speech and Expression

The Online Safety Act 2023 has ignited significant debate regarding its potential impact on freedom of speech and expression online, with both proponents and critics presenting strong arguments.

Potential Drawbacks and Criticisms:
Critics, including politicians, academics, journalists, and human rights organizations, contend that the Act poses a significant threat to the right to privacy and freedom of speech and expression.2 Organizations such as Article 19 have characterized the Act as "an extremely complex and incoherent piece of legislation" that "fails in effectively addressing the threat to human rights" like freedom of expression and access to information.2 Similarly, Big Brother Watch and the Open Rights Group have labeled it a "censor's charter" that undermines freedom of expression and privacy.2 These groups express concern that the Act grants the UK government extensive powers to regulate speech, set enforcement priorities, and pressure platforms into removing content without sufficient judicial oversight.2 Civil liberties organizations specifically criticize proposals to restrain the publication of lawful speech deemed

harmful, fearing this could lead to censorship or the "silencing of marginalised voices and unpopular views".2 This suggests concerns that the legislation might lead to over-censorship or a chilling effect on legitimate online discourse, particularly given its focus on regulating content that amounts to a criminal offense and a wider category of "harmful" content for child users.3
A major point of contention is the Act's implications for end-to-end encryption. The Act requires platforms, including end-to-end encrypted messengers, to scan for child pornography.[2] Cybersecurity experts argue that this requirement is technically impossible to implement without undermining users' privacy.[2] While the government has stated it does not intend to enforce this provision until it becomes "technically feasible," the powers allowing Ofcom to mandate the weakening of end-to-end encryption were not removed from the Act, meaning Ofcom can issue such notices at any time.[2] This has led to strong concerns from major technology firms; Apple Inc. called the legislation a "serious threat" to end-to-end encryption, and Meta Platforms stated it would rather have WhatsApp and Facebook Messenger blocked in the UK than weaken encryption standards.[2] The European Court of Human Rights has ruled that requiring degraded end-to-end encryption "cannot be regarded as necessary in a democratic society" and is incompatible with Article 6 of the European Convention on Human Rights.[2]

The Act is also criticized for potentially burdening companies with "disproportionate obligations".[3] This could indirectly impact freedom of expression if companies, in an effort to comply and avoid severe penalties (up to £18 million or 10% of annual worldwide revenue), become overly cautious in what content they allow, leading to the removal of borderline or controversial but legal speech.[3] Some websites have already announced their closure or blocked UK users due to the high cost of legal compliance or to avoid penalties under the Act. Examples include London Fixed Gear and Single Speed, Microcosm, Gab, and Civit.ai.[2] Following the enactment, there was a significant rise in downloads of VPN services by UK users, as these can circumvent age verification requirements by routing traffic through other countries.[2] A petition calling for the repeal of the law attracted over 500,000 signatures on the UK Parliament petitions website.[2]

Potential Benefits and Support:
Supporters of the Act, including Prime Minister Sir Keir Starmer, the National Crime Agency, and the National Society for the Prevention of Cruelty to Children (NSPCC), emphasize its necessity for child protection from online harms.2 The NSPCC called its passage "a momentous day for children" that would help prevent abuse.2 The Act creates a new duty of care for online platforms, requiring them to take action against illegal content and legal content that could be harmful to children where children are likely to access it.2 Platforms failing this duty are liable to significant fines.2 The government has heralded the Act as a development that will "make the UK the safest place in the world to be online".3
The Act obliges technology platforms to introduce systems that will allow users to better filter

out harmful content they do not want to see.[2] It also adds new offenses to the Sexual Offences Act 2003, such as sending images of a person's genitals (cyberflashing) and sharing or threatening to share intimate images.[2] It updates and extends existing communication offenses, including the false communications offense and creating offenses for sending messages conveying threats of death or serious harm, and sending or showing flashing images electronically with intent to cause harm to individuals with epilepsy ("epilepsy trolling").[2] It also criminalizes encouraging or assisting serious self-harm.[2] These provisions aim to curb specific harmful online behaviors.

The Act includes provisions allowing eligible entities to bring super-complaints on behalf of consumers, with the process for doing so set out in regulations in July 2025.[2] Furthermore, it has provisions to impose legal requirements ensuring that content removals do not arbitrarily remove or infringe access to journalistic content.[2] Large social networks are also required to protect "democratically important" content, such as user-submitted posts supporting or opposing political parties or policies.[2]

4.2 Implications for User Privacy and Data Protection

The Online Safety Act 2023 has several implications for user privacy and data protection, data collection, sharing, and user rights, particularly concerning age verification and content scanning.

Implications for User Privacy and Data Protection:
The requirement for platforms, including end-to-end encrypted messengers, to scan for child pornography is a significant concern.2 Experts argue that this is not possible to implement without undermining users' privacy.2 While the government has stated it does not intend to enforce this provision until it becomes "technically feasible" 2, the underlying powers remain in the Act. This creates a tension for technology firms that prioritize user privacy through strong encryption. Civil liberties organizations like Big Brother Watch and the Open Rights Group have criticized the bill for its proposals, stating they pose a threat to the right to privacy.2 A cybersecurity expert described the Act's surveillance provisions as "technically dangerous and ethically questionable," suggesting it could make the internet less safe and lead to mass surveillance.2 Major technology firms, including Apple Inc. and Meta Platforms, have raised concerns that the Act could force them to weaken security features designed to protect users from surveillance and end-to-end encryption, with some stating they would rather be blocked in the UK than weaken encryption standards.2
Despite these concerns, the Act does include "A duty to have regard to the importance of...

protecting users from unwarranted infringements of privacy, when deciding on, and implementing, safety policies and procedures".[2] This provision was intended to address privacy concerns, particularly those related to age verification requirements.

Implications for Data Collection and Sharing:
Section 12 of the Act mandates that service providers use age verification or age estimation technology to prevent children from accessing "primary priority content that is harmful to children," including pornographic images and content encouraging self-harm or suicide.2 This applies to all services that make such content available, including social networks and internet pornography services.2 This implies a need for companies to collect and process age-related data, which has direct implications for user privacy and data protection. The Wikimedia Foundation and Wikimedia UK have rejected calls to implement age verification or identity checks, citing concerns about data minimisation and privacy.2 Providers of Category 1 and Category 2A services are also required to publish information about their illegal content risk assessments and, where applicable, children's risk assessments, and supply full records to Ofcom.3 This increased transparency, while not directly about user data, could indirectly impact user privacy by revealing more about how platforms identify and manage risks related to user-generated content.
Implications for User Rights:
Beyond privacy, critics argue that the Act poses a threat to the right to freedom of speech and expression, potentially leading to censorship or the "silencing of marginalised voices and unpopular views".2 Conversely, the Act includes provisions to ensure that content removals do not arbitrarily remove or infringe access to journalistic content.2 Large social networks are also required to protect "democratically important" content, such as user-submitted posts on political parties or policies.2 The Act obliges technology platforms to introduce systems that will allow users to better filter out harmful content they do not want to see.2 The Act also contains provisions allowing eligible entities to bring "super-complaints" on behalf of consumers, with the process for this set out in regulations in July 2025.2 Some websites, including forums and non-commercial hosting providers, have announced their closure or blocked UK users due to the high cost of legal compliance and concerns about the Act's implications.2 One academic suggests that while the Act provides a legal basis for removing illegal content, it does not adequately protect users from disinformation and online harassment, which is necessary for ensuring political participation.2

4.3 Operational and Financial Impact on Online Service Providers and Tech Companies

The Online Safety Act 2023 will impose extensive new regulatory requirements on an estimated

100,000 companies offering user-to-user services or search services in the UK.[3] This signifies a substantial operational and financial impact across the industry.

Compliance Requirements and Deadlines:
Companies are required to assess and manage safety risks arising from content and conduct on their services.3 This involves carrying out illegal content risk assessments to consider the likelihood of specific types of illegal content being available and the risk of service misuse for offenses.3 For services likely to be accessed by children, a separate children's risk assessment is required, focusing on online harms children of different age groups could face.3
Key compliance deadlines include:

- **March 31, 2025:** Deadline for completing illegal harms risk assessments and implementing mitigation controls, following guidance published in December 2024.[6]
- **April 16, 2025:** Deadline for Child Access Assessments, published in January 2025, determining if children can and are likely to access the platform.[6]
- **July 24, 2025:** Deadline for Children's Risk Assessments, published in April 2025, mandatory if the Child Access Assessment indicates a need.[6]
- Guidance on the protection of Women and Girls was issued in March 2025, addressing misogyny, harassment, and intimate image abuse.[6]

Companies will need to implement mitigation measures tailored to the identified risks. These can include regulatory compliance frameworks, design adjustments to functionalities and algorithms, terms of use policies, user access controls, content moderation, user support, and staff training.[3] The Act does not contain explicit exceptions for small and micro enterprises, meaning "all in-scope user-to-user and search services, large or small, will need to take action to comply with the new duties".[3] This broad application indicates a significant operational impact across the industry, potentially disproportionately affecting smaller entities. Developing the required risk assessments can be a "time and resource intensive exercise," necessitating organizations to evaluate their internal "capacity, knowledge and expertise" or seek expert support.[6]

Potential Penalties for Non-Compliance:
Ofcom has several robust enforcement powers:

- **Fines:** Up to £18 million or 10% of qualifying world revenue, whichever is greater.[2] This significantly exceeds the maximum penalty under the EU's Digital Services Act (6% of total worldwide annual turnover).[3]
- **Business Disruption:** A new power allows Ofcom to order a third party (e.g., payment providers, ad networks) to cease doing business with the offending provider.[6]
- **Information Gathering Powers:** Ofcom can request documents, conduct audits, and interview staff members.[3]
- **Criminal Liability for Senior Managers:** Senior managers may also face potential

criminal liability in certain situations, including if they fail to ensure their organization follows information requests from Ofcom or provide false information during interviews.[3]

- **Reputational Risk:** Non-compliance can also lead to adverse media coverage, resulting in significant reputational damage.[6]

Ofcom will adopt a risk-based and proportionate enforcement approach, prioritizing larger organizations and "big name" platforms due to their wider reach and higher risk.[6] While larger providers have already been contacted, all organizations covered by the Act must comply, and smaller organizations are not exempt, especially in cases of serious breaches.[6] Companies are advised to prepare for phased implementation by gathering user data (especially age groups) and reviewing current risk management activities.[3] Well-drafted and implemented terms of service will also be crucial for compliance.[3]

4.4 Mechanisms for Enforcement and Accountability

The Online Safety Act 2023 empowers Ofcom, the national communications regulator, to enforce its provisions.[2] Ofcom adopts a risk-based and proportionate approach to enforcement, focusing efforts on larger organizations and "big name" platforms due to their greater reach and higher risk.[6]

Enforcement Mechanisms:

- **Blocking Access:** Ofcom can block access to specific user-to-user services or search engines within the United Kingdom. This includes interventions by internet access providers and app stores.[2]
- **Service Restriction Orders:** The regulator can impose requirements on ancillary services that facilitate the provision of regulated services through "service restriction orders".[2] Examples include services that enable fund transfers, search engines that display or promote content in search results, and services that facilitate the display of advertising on a regulated service.[2]
- **Court Application:** Ofcom must apply to a court for both Access Restriction and Service Restriction Orders.[2]
- **Secretary of State's Power to Direct Ofcom:** Section 44 of the Act grants the Secretary of State the power to direct Ofcom to modify a draft code of practice for online safety. This power can be exercised for reasons of public policy, national security, or public safety.[2] Ofcom is obligated to comply with such directions and submit a revised draft to the Secretary of State. The Secretary of State can issue further directions for modifications and,

once satisfied, must lay the modified draft before Parliament. Additionally, the Secretary of State can remove or obscure information before presenting the review statement to Parliament.[2]

- **Super-Complaints:** The Act includes provisions allowing eligible entities to bring "super-complaints" on behalf of consumers. The process for these super-complaints was established by regulations in July 2025.[2]

Roles of Regulatory Bodies:

- **Ofcom:** As the national communications regulator, Ofcom is the primary body responsible for enforcing the Online Safety Act 2023. Their powers include blocking access to services, imposing service restriction orders, and developing codes of practice for online safety.[2]
- **Secretary of State for Science, Innovation and Technology:** This government official introduces the Act in the Commons and holds the power to direct Ofcom on matters of public policy, national security, or public safety related to online safety codes of practice.[2]
- **Lord Parkinson of Whitley Bay, Parliamentary Under-Secretary of State for Arts and Heritage:** This official introduces the Act in the Lords.[2]

Legal Challenges:
The Act has faced several legal challenges and significant criticism. The Wikimedia Foundation launched a legal challenge in May 2025 against potential designation as a "category one" service under the Act.2 This designation would subject Wikipedia to the most stringent requirements, which the Foundation argued would compromise Wikipedia's open editing model and invite state-driven censorship or manipulation.2 While
The Daily Telegraph reported in July 2025 that Wikipedia might restrict access for UK users if full compliance is insisted upon, the High Court of Justice dismissed this challenge in August 2025.[2]

The provisions concerning end-to-end encryption have also been a focal point of legal and industry concerns. The Act requires platforms, including end-to-end encrypted messengers, to scan for child pornography.[2] Experts argue that this is not possible without undermining user privacy.[2] Despite the government's statement that it does not intend to enforce this provision until it becomes "technically feasible," the provisions allowing Ofcom to require the breaking of end-to-end encryption technology were not removed from the Act, and Ofcom can issue such notices at any time.[2] Major technology firms like Apple Inc. and Meta Platforms have raised concerns, with Meta stating it would rather have WhatsApp and Facebook Messenger blocked in the UK than weaken encryption standards.[2] The European Court of Human Rights ruled in February 2024 that requiring degraded end-to-end encryption "cannot be regarded as necessary in a democratic society" and was incompatible with Article 6 of the European Convention on Human Rights.[2]

4.5 Public and Expert Reactions, Criticisms, and Support

The Online Safety Act 2023 has elicited a wide range of reactions from the public, academics, charities, human rights organizations, and industry, leading to ongoing debates about its implications.[2]

Public Responses:
Following the enactment of the law, there was a notable increase in downloads of VPN services by UK users, as these can bypass age verification requirements by routing traffic through countries without such regulations.2 Some users also successfully circumvented photo-based age verification services, like Persona, by using images of characters from video games.2 A petition calling for the repeal of the law garnered over 500,000 signatures on the UK Parliament petitions website.2
Academic Responses:
Academics have expressed mixed views. Alan Woodward, a cybersecurity expert, characterized the Act's surveillance provisions as "technically dangerous and ethically questionable," arguing that the government's approach could make the internet less safe and that the Act makes mass surveillance "almost an inevitability".2 Elena Abrusci of Brunel Law School suggested that while the Act provides a legal basis for removing illegal content, it inadequately protects users from disinformation and online harassment, which are crucial for political participation.2
Charities and Human Rights Organizations:
Support for the Act primarily comes from child protection advocates. The National Society for the Prevention of Cruelty to Children (NSPCC) hailed the Act's passage as "a momentous day for children," stating it would help prevent abuse.2 The Samaritans cautiously supported the final Act, viewing it as a step forward but criticizing the government for not fully achieving its ambition of making the UK the "safest place to be online".2
However, human rights organizations have been highly critical. British human rights organization Article 19 warned that the Act is "an extremely complex and incoherent piece of legislation" and "fails in effectively addressing the threat to human rights" such as freedom of expression and access to information.[2] Mark Johnson of Big Brother Watch labeled it a "censor's charter" that undermines the right to freedom of expression and privacy.[2] The European Court of Human Rights' ruling against degraded end-to-end encryption further underscores these concerns.[2]

Industry Responses:
The industry response has been largely critical, driven by concerns over compliance costs and technical feasibility. Several websites announced their closure due to the Act, citing high legal

compliance costs, including London Fixed Gear and Single Speed and Microcosm.2 Some sites, such as the alt-tech social networking service Gab and Gen AI platform Civit.ai, blocked UK users to avoid penalties.2 Major technology firms expressed concerns about user privacy and encryption; Apple Inc. called the legislation a "serious threat" to end-to-end encryption, and Meta Platforms stated it would rather have WhatsApp and Facebook Messenger blocked in the UK than compromise encryption standards.2

Despite the criticisms, some websites and apps have begun introducing age verification systems for UK users in response to Ofcom's deadlines. These include pornographic websites, as well as social networks like Bluesky, Discord, X, dating apps such as Tinder, Bumble, Feeld, Grindr, and Hinge, and streaming services like Spotify. Reddit also implemented age verification for adult content for its UK users.[2]

Political Responses:

Politically, the Act has seen both strong support and strong opposition. Prime Minister Sir Keir Starmer stated the Act is necessary to protect children from online content like "suicide sites".2 The National Crime Agency also emphasized the legislation's necessity for safeguarding children from online harms.2 In contrast, Reform UK leader Nigel Farage called the Act "borderline dystopian" and vowed to repeal it if elected, with fellow Reform leader Zia Yusuf describing the legislation as "an assault on freedom".2 The government of Jersey notably decided not to enforce the law in its territory, citing "inadequacies" and opting to develop its own online safety legislation instead.2

Ongoing Debates:

The Act continues to be a subject of intense debate, particularly concerning its impact on freedom of speech, privacy, and the feasibility of its technical requirements. The controversial powers allowing Ofcom to potentially break end-to-end encryption remain in the Act, despite a ministerial statement that they would not be used immediately, leading to ongoing concerns from tech firms about withdrawing from the UK market.2 The Wikimedia Foundation and Wikimedia UK have strongly objected to the Act, arguing it risks undermining non-profit and community-governed websites like Wikipedia.2 They have rejected implementing age verification or identity checks due to concerns about data minimization, privacy, and editorial independence, and in June 2023, urged lawmakers to exempt public interest platforms.2 The Wikimedia Foundation's legal challenge against potential designation as a "category one" service, fearing it would compromise Wikipedia's open editing model and invite state-driven censorship, was ultimately dismissed in August 2025.2

4.6 Comparison with Other Online Safety Laws (e.g., EU Digital Services Act)

The Online Safety Act (OSA) and the EU's Digital Services Act (DSA) both aim to regulate online service providers to create safer digital spaces, particularly by mitigating risks of harm arising from illegal and harmful content.[7] While sharing similar overarching goals, they adopt markedly different approaches in scope, specificity, and the obligations imposed on digital platforms.[7]

Scope and Focus:

The DSA takes a broader approach, covering a wide range of issues including intellectual property (IP) infringement, illegal goods, dark patterns, and crisis response.7 In contrast, the OSA adopts a more laser-focused, detailed approach, primarily dealing with illegal content.7 However, within this specific focus, the OSA is a significantly more detailed and granular piece of legislation.7 The OSA specifically targets content that constitutes a *criminal offense* under UK law, whereas the DSA addresses all forms of "illegal" and infringing content, including civil law infringements like defamation or intellectual property rights.[3]

Definitions of Illegal Content:

The DSA broadly defines illegal content, encompassing online actions that are illegal under member state laws, EU treaties, and EU-wide legislation, as well as content or user behavior that was previously unlawful offline but has now transitioned to the online sphere.7 The OSA, conversely, delineates two categories of illegal content—general offenses and "priority offenses," such as child sexual exploitation and abuse and terrorist content—providing in-depth specifications for defining what qualifies as illegal content.3

Platform Obligations:

A key distinction lies in the approach to platform obligations. The DSA primarily functions under a "notice and takedown" complaint system, typically using form-based procedures.7 This is a more reactive model. In contrast, the OSA requires platforms to actively monitor content, departing from the traditional voluntary approach, with the explicit goal of preventing users from encountering clearly illegal content.7 It explicitly states that, when technically feasible, companies must employ algorithms or other product features for this proactive purpose.7 This represents a fundamental shift towards proactive measures in the UK.

Classification of Platforms and Exemptions:

Under the DSA, services attain the titles of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) when they have or exceed 45 million monthly users, with lower compliance thresholds for smaller companies.7 Conversely, the OSA has not provided explicit details on its approach to assessing company categories other than it will rely on risk evaluations and the size of the UK user base.7 Crucially, the OSA does not create exemptions in its compliance framework based on a company's size, in contrast to the DSA.3 Ofcom has stated that "all in-scope user-to-user and search services, large or small, will need to take action to comply with the new duties".3

Protection of Minors:
Both the DSA and OSA emphasize the importance of protective measures for minors, albeit through different approaches.7 The UK's Online Safety Act goes beyond just restricting harmful content by requiring platforms to be designed with children in mind—offering clear reporting tools, easy-to-understand terms of service, and real support when problems arise.8 It also mandates the use of age assurance tools for services likely to be accessed by children.8
Comparison with US Legislation:
While the UK is fully enforcing its Online Safety Act, the US is considering less ambitious laws.8 For instance, the Kids Online Safety Act (KOSA) in the US, while aiming to shield minors, has faced criticism from civil liberties groups like the ACLU, who argue it would threaten privacy, limit minors' access to vital resources, and restrict adults' freedom of expression.8 KOSA would require "the strongest privacy settings for kids by default" and give parents "new controls," along with requiring platforms to "prevent and mitigate specific dangers to minors" like promotion of suicide, eating disorders, and sexual exploitation.8 However, unlike the UK's broad proactive duties, the US approach tends to be more narrowly focused on specific harms to children or reactive measures. Australia also passed a law banning anyone under 16 from accessing social media.8 The UK's Act stands out for its comprehensive, proactive duty of care framework and its broad extraterritorial reach compared to many international counterparts.

5. Conclusions

The Online Safety Act 2023 marks a profound shift in the United Kingdom's approach to regulating online content, moving from a reactive "notice and takedown" model to a proactive "duty of care" framework. This legislative evolution is a direct response to widespread concerns regarding online harms, particularly those affecting children, and aims to establish the UK as a global leader in online safety.

The Act's broad and extraterritorial scope, encompassing a wide array of user-to-user services, search engines, and pornography providers regardless of their physical location, signals an intent to ensure universal accountability for online safety within the UK's digital sphere. A key distinction is the absence of explicit exemptions for small and micro enterprises, which, while aiming for a consistent baseline of safety, places a significant operational and financial burden across the industry. The Act's primary focus on content that constitutes a criminal offense for adults, while extending to a broader category of "harmful" content for children, reflects a nuanced regulatory strategy. For adults, the emphasis shifts towards empowering users with greater control over their content exposure.

The implementation of the duty of care framework necessitates that online platforms transform

into sophisticated risk management entities. This requires substantial investment in algorithmic accountability, systemic design changes, and robust internal compliance mechanisms. Ofcom, as the designated "super-regulator," is endowed with unprecedented enforcement powers, including substantial fines and the ability to disrupt business operations through service restriction orders, making compliance an existential imperative for platforms. The phased implementation of the Act, with staggered deadlines for risk assessments and mitigation measures, aims to facilitate industry adaptation, though the technical challenges, particularly concerning end-to-end encryption, remain a contentious issue.

The Act introduces new criminal offenses, directly targeting specific online harms like "epilepsy trolling," encouraging self-harm, and cyberflashing. This expansion of criminal law into the digital realm reflects a legislative effort to hold individuals directly accountable for egregious online conduct, thereby integrating online platforms more deeply into the broader law enforcement system.

Public and expert reactions to the Act have been polarized. Supporters laud its potential for child protection and its proactive approach to online harms. Critics, however, voice significant concerns regarding its potential to curtail freedom of speech and expression, undermine user privacy, and impose disproportionate compliance burdens that could stifle innovation or lead to service withdrawals from the UK market. The ongoing debates, particularly around the feasibility of encryption scanning and the potential for executive influence over regulatory functions, highlight the complex balance the Act attempts to strike between safety and fundamental rights.

In conclusion, the Online Safety Act 2023 represents a landmark attempt to regulate the digital landscape, setting a precedent for proactive online safety governance. Its comprehensive scope, stringent duties, and formidable enforcement mechanisms are poised to fundamentally reshape the operations of online service providers. While lauded for its child protection focus and proactive stance against illegal content, its long-term impact on freedom of expression, user privacy, and the competitive landscape for smaller online businesses remains a subject of ongoing scrutiny and adaptation. The Act's success will ultimately depend on Ofcom's pragmatic and proportionate enforcement, as well as the industry's capacity to innovate and adapt to these new, demanding regulatory realities.

Works cited

1. en.wikipedia.org, accessed on August 16, 2025, https://en.wikipedia.org/wiki/Online_Safety_Act_2023#:~:text=The%20Online%20Safety%20Act%202023,Kingdom%20to%20regulate%20online%20content.
2. Online Safety Act 2023 - Wikipedia, accessed on August 16, 2025, https://en.wikipedia.org/wiki/Online_Safety_Act_2023

3. ONLINE SAFETY ACT 2023 - Wilson Sonsini, accessed on August 16, 2025, https://www.wsgr.com/a/web/vExo8JtW6yDSp7K3qt6qzX/online-safety-act.pdf
4. Final Statement of Strategic Priorities for Online Safety - GOV.UK, accessed on August 16, 2025, https://www.gov.uk/government/publications/statement-of-strategic-priorities-for-online-safety/final-statement-of-strategic-priorities-for-online-safety
5. UK Online Safety Act 2023 - Latham & Watkins LLP, accessed on August 16, 2025, https://www.lw.com/admin/upload/SiteAttachments/UK-Online-Safety-Act-2023.pdf
6. UK's Online Safety Act 2023: What You Need to Know - BDO, accessed on August 16, 2025, https://www.bdo.co.uk/en-gb/insights/advisory/risk-and-advisory-services/uks-online-safety-act-2023-what-you-need-to-know
7. The Differences Between the Online Safety Act & the Digital Services Act | TrustLab Blog, accessed on August 16, 2025, https://www.trustlab.com/post/the-differences-between-the-online-safety-act-the-digital-services-act
8. UK Online Safety Act Tougher Than Proposed US Law - ConnectSafely, accessed on August 16, 2025, https://connectsafely.org/uk-enforcing-tough-online-protection-law-as-us-considers-less-ambitious-law/
9. Effective enforcement of the Online Safety Act and Digital Services Act: unpacking the compliance and enforcement regimes of the UK and EU's online safety legislation - Taylor & Francis Online, accessed on August 16, 2025, https://www.tandfonline.com/doi/abs/10.1080/17577632.2025.2459441