ISD | Institute for Strategic Dialogue

ALFRED LANDECKER FOUNDATION

Policy brief

Addressing Illegal Harms on Small and Emerging Platforms: Regulatory Challenges and Gaps

Arthur Bradley

About the Digital Policy Lab

The Digital Policy Lab (DPL) is an inter-governmental working group focused on charting a policy path to prevent and counter the spread of influence operations, hate speech, extremist and terrorist content online. It is convened by the Institute for Strategic Dialogue (ISD) and consists of representatives from relevant ministries and regulatory bodies from selected liberal democracies. The DPL aims to foster inter-governmental exchange, provide policymakers and regulators with access to sector-leading expertise and research, and build an international community of practice around key challenges in the digital policy space.

About this paper

As part of the DPL, ISD organised two working group meetings on the topic of regulating emerging platforms and technologies in May 2025. The working group consisted of DPL members including regulators, competent authorities and law enforcement from multiple jurisdictions, and representatives from civil society and academia. While participants' contributions have informed the analysis in this paper, the views expressed within do not necessarily reflect the views of all participants, nor any governments involved in this project.

Acknowledgements

We would like to thank all members and participants of the working group for their contributions. In particular, we would like to give thanks to the speakers in the two sessions, namely Dr. Ali Fisher (Università Cattolica del Sacro Cuore Milano and Human Cognition Limited), Léa Ronzaud (researcher), Sean McCafferty (Metropolitan University Prague and VOX-Pol Institute), Lucile Petit (Arcom, France), Antonis Samouris (Europol), Maygane Janin (Christchurch Call Foundation) and representatives from Ofcom (UK). Additionally, we would like to thank participants from the following ministries and regulatory authorities: Canadian Heritage (Canada), Public Safety Canada (Canada), the Department of Tourism, Culture, Arts, Gaeltacht, Sports and Media (Ireland), Authority for Consumers and Markets (Netherlands), Ministry of Justice and Security (Netherlands), and Council for Media Services (Slovakia). We also thank the Center for the Study of Democracy for their participation in the working group.

About the author

Arthur Bradley is an independent specialist in tracking and analysing terrorist and other malevolent use of online platforms. He provides consultancy services to the private sector, non-governmental organisations, intergovernmental organizations, academia, public sector institutions and law enforcement agencies. Arthur was previously open-source intelligence (OSINT) manager at Tech Against Terrorism.

ALFRED LANDECKER FOUNDATION

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2025). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address 3rd Floor, 45 Albemarle Street, Mayfair, London, W1S 4JL. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved.

www.isdglobal.org

Contents

Executive Summary

This policy brief explores the challenges faced by regulators when working to facilitate improvements in online safety across numerous small but high-risk platforms and emerging technologies. It includes an introduction to the threat landscape on these platforms and services, with a particular focus on illegal harms such as terrorism, violent extremism and illegal hate speech[1], while also recognising the potential presence of other serious illegal harms such as child sexual abuse material (CSAM). It also includes an assessment of key gaps in effective responses to these threats by regulators, policy makers, law enforcement, civil society and companies themselves. It considers key online safety regulations adopted in recent years particularly the UK Online Safety Act (OSA) and the EU's Digital Services Act (DSA) and Terrorist Content Online Regulation (TCO). Australia's Online Safety Act (OSA) is also considered.

Key Findings

- While regulation such as the DSA or UK OSA tend to prioritise larger platforms on the basis of their greater numbers of users, a significant quantity of illegal harms are concentrated on a broad ecosystem of smaller or emerging platforms and technologies. Some of these platforms present themselves as being bastions of free speech and are ideologically opposed to the requirements of online regulation. For others, there is a comparative lack of capacity, or occasionally willingness, to respond to these threats.

- Illegal content ecosystems such as terrorist networks tend to be highly adaptive and agile in their use of multiple platforms, including platform migrations in response to regulatory actions. This issue is compounded by difficulties experienced by some regulators in gaining access to real-time primary data, which could be used to inform targeted and proportionate enforcement.

- Current regulatory frameworks are insufficiently adapted to tackling high-risk small platforms. Regulatory regimes like the EU DSA and UK OSA have reactive elements which require platforms to remove illegal content when it is reported to them. Under Australia's OSA, online safety regulator eSafety can facilitate the removal of serious online abuse, illegal and restricted online content from platforms once it is reported. However, these frameworks generally do not apply risk-based approaches to small high-risk platforms in the way that they do for larger online services that reach regulatory user number thresholds.

- Challenges in identifying, assessing and enforcing illegal and harmful content across numerous small but high-risk platforms and services are hindering a more effective response in these spaces. Despite the emergence of coordination mechanisms such as the European Board for Digital Services or the Global Online Safety Regulators Network, international cooperation and coordination could still be further enhanced.[2] Additionally, a lack of regulator access to primary data on illegal harms ecosystems can also compound these issues.

Introduction

Many online regulatory regimes including the EU's Digital Services Act (DSA) and the UK's Online Safety Act (OSA) consider the size of a platform's userbases as a decisive rationale to impose the heaviest obligations on them. However, a broad ecosystem of smaller platforms exists whose features and usage by malicious actors also pose significant risks to online safety. According to Ofcom in January 2025, "more than 100,000" online services were likely to be in the scope of the OSA in the UK alone, many of which can be categorised as small or emerging platforms.[3]

A significant body of research in recent years has evidenced how terrorist, extremist and other malicious actors proliferate across and heavily exploit a significant number of such small or emerging services. These actors still use mainstream platforms but their decision to adopt particular small services is a tactical choice: often this reflects companies' lack of capacity or willingness to identify, verify and remove illegal content.[4] Terrorist networks and other malicious actors are also adept at migrating between platforms or using multiple services simultaneously to evade detection or circumvent the removal of their accounts and material.

However, the use of smaller or emerging platforms is not just a matter of opportunism but a structural feature of how extremist and terrorist ecosystems operate. Dr Ali Fisher describes decentralised and highly adaptive networks affiliated with the Islamic State (IS) group, where content production and dissemination are distributed across a range of loosely-connected actors. Large platforms are often used for amplification, recruitment and to help propaganda reach broader audiences; more niche services function as operational hubs including for coordination, ideological consolidation and content storage.[5]

It can be difficult for regulators, law enforcement and other online safety practitioners to identify, prioritise and engage the platforms that pose a significant risk, regardless of their size, particularly without access to real-time primary data.[6] Furthermore, despite the often-cited desire for regulation to be future proof, its applicability to emerging technologies is also sometimes unclear for both online safety practitioners and the developers of those technologies.[7]

This policy brief provides insight into the challenges faced by regulators, law enforcement, technology companies and other online safety practitioners when tackling illegal activity on small or emerging platforms that pose significant risk of harm. It outlines the current online threat landscape on smaller or emerging platforms, and analyses blockers to a more effective response. This includes an analysis of the most recent and comprehensive online regulation.

To highlight a possible approach to prioritising and deconflicting engagement with such platforms, this brief also includes a sample set of small platforms that have been most heavily exploited by Islamist terrorist networks between April 2024 and April 2025, based on primary data, mapped according to the jurisdiction in which they are registered. Finally, this brief identifies policy implications and provides recommendations for a more streamlined and effective response to these issues.

What is the threat?

Cross-platform ecosystems

A significant body of research in recent years has shown how illegal harms (including terrorism, illegal hate speech and CSAM)[8] proliferate across a significant number of online services. These harms occur not only on the largest platforms subject to the most stringent regulatory obligations, but also across smaller and less regulated services that perpetrators often use in parallel. In the case of terrorist and violent extremist activity online, content manifests in wide-reaching cross-platform networks that often congregate on smaller or emerging services. Criminals sharing CSAM, for example, also use smaller file-sharing platforms, dedicated websites and messaging apps with varying degrees of encryption to share and host illegal content in tandem with their use of mainstream services.[9] These actors choose smaller platforms for a variety of reasons including features, marketing and a perception that their content is less likely to be removed than on larger services.

There are also multiple examples of the ways in which malicious actors have exploited emerging technologies. White supremacists and other groups promoting violent extremist content have heavily used alternative social media platforms ('alt-tech') in recent years. These include decentralised platforms and the Federated Universe (Fediverse), a decentralised network of independent social media servers ("instances") that communicate via a shared protocol like ActivityPub. While this approach enables interoperability and user control, its open and self-governed structure has also led extremist communities to establish their own loosely connected platforms without external accountability.[10] Extremist actors have also exploited gaming or game-adjacent services for communication and radicalisation purposes such as modifying sandbox games to create simulations of real-world mass shootings.[11]

Shifting threat landscapes

While the overall userbase of these services is small compared to mainstream platforms, smaller platforms are frequently the first venue for illegal content before it spreads to mainstream social media. Despite often having a comparatively small userbase, these services are regularly relied on by terrorist, extremist and other malicious actors as more stable alternatives to mainstream platforms; these actors direct users of mainstream services to content and online spaces there via URLs shared on social media. As a result, traffic to these kinds of services can often shift rapidly over a short period of time.

Following a livestreamed white supremacist terrorist attack in Buffalo, New York in May 2022, a copy of the attack footage gained more than 3 million views on video-sharing platform Streamable after the original livestream was removed by Twitch.[12] A link to the footage on Streamable had been shared more than 47,000 times on Facebook.[13] Following the incident in Buffalo, Streamable has since become a member of the Global Internet Forum to Counter Terrorism (GIFCT), giving it access to a broader range of tools to counter similar incidents in future.[14]

There are also several examples in which large numbers of users may arrive on a new platform, particularly when blunt action is taken against whole platforms. Following the 6 January 2021 attack on the US Capitol, Amazon Web Services (AWS) suspended services to Parler, which had 12.3 million monthly users at the time.[15] Among the platforms some users migrated to was Telegram, where administrators of white supremacist channels traded playbooks on how to radicalise the new Trump-supporting arrivals to neo-Nazism.[16]

Following an announcement by the US government in January 2025 that TikTok would be banned in the US, both neo-Nazi and white supremacist extremists and networks affiliated with IS group established accounts on XiaoHongShu (aka "RedNote"), a Chinese platform with similar features.[17] The announcement led to around 700,000 new users joining over the course of just two days, many of them American.[18]

Evidencing these rapidly shifting ecosystems across numerous smaller or emerging platforms and services requires dedicated and specialist resources that can be lacking for some regulators or other public sector online safety practitioners. However, there are several non-governmental organisations and companies whose work is dedicated to this monitoring and analysis. This provides an opportunity for collaboration and coordination to offer government agencies access to real-time, primary data, informing effective and targeted regulatory activity based on the evolving threat picture.

This brief includes a list of the top 20 platforms used by IS groups and Al-Qaeda networks over the past year (based on the total number of outlinks captured from core IS group and Al-Qaeda channels across platforms between April 2024–April 2025); the aim of this list is to illustrate a framework for platform prioritisation based on threat, and to allocate those companies to their relevant regulator or competent authority, based on the jurisdiction in which they are registered.

Regulatory approaches to small but high-risk platforms

Some regulatory regimes have tiered obligations for platforms based either partially or entirely on their number of monthly users. As a result, in some cases smaller platforms that still pose high levels of risk are not subject to the strongest obligations. There is a risk that this could lead to gaps in effective responses to the spread of illegal or otherwise harmful content on these services, its algorithmic amplification, as well as a lack of transparency requirements. To illustrate how these regimes apply to small or emerging platforms and technologies, the table below provides a brief overview of key regulation in the EU, UK, and Australia.

EU Digital Services Act (DSA)

The DSA applies proportional measures to online intermediary services (dependent on a service's type or size) to prevent illegal and harmful activities online and ensure user safety. It came into force for all platforms operating in the EU in February 2024, although its rules had applied to the largest services – described as Very Large Online Platforms (VLOPs), Very Large Online Search Engines (VLOSEs) and collectively Very Large Online Platforms and Search Engines (VLOPSEs) since August 2023.[19] VLOPSEs are defined in the regulation as services that reach more than 10% of the 450 million consumers in the EU.[20] For these services, the DSA introduces heightened obligations to identify, assess and mitigate "systemic risks" including the dissemination of illegal content, negative effects on fundamental rights, risks to electoral processes, gender-based violence, and harms to public security or to minors.

Online platforms and hosting services that do not meet the VLOPSE threshold are subject to a comparatively reduced range of obligations. These include provisions on the removal of illegal content, notice and action mechanisms, enhanced transparency reporting, and child protection measures.[21]

So-called "micro and small enterprises" are exempt from certain obligations under the DSA; these are defined as companies with a staff headcount of less than 50 and a turnover of less than EUR 10 million.[22] These exemptions are in relation to duties such as transparency mechanisms, user appeal functions and trusted flagger programmes, rather than the core obligations to remove illegal content once it is reported.[23]

EU Terrorist Content Online (TCO) regulation

The Terrorist Content Online (TCO) has been applicable in the EU since June 2022. It aims to ensure that hosting service providers remove terrorist content online within an hour of receiving removal orders from Member States' competent authorities. The regulation applies to Hosting Service Providers (HSPs) that offer services in the EU, regardless of whether they are headquartered there.[24]

Under the TCO, online platforms must take proactive measures when they are considered to be "exposed" to terrorist content. Such rulings come when a network is subjected to at least two removal orders from competent authorities in a period of 12 months.[25] The measures they take are contingent on the assessed level of exposure, their resources, and their size.[26] For example, the Irish regulator Coimsiún na Meán ordered X (formerly Twitter), TikTok and Instagram in November 2024 to take "necessary measures" after determining that these platforms had been exposed to terrorist content.[27]

While the TCO applies to HSPs of all sizes, the specific measures platforms are required to implement to address the misuse of their services, as well as the penalties for non-compliance, may be adjusted based on the classification of these platforms as a start-up, micro-, small-, or medium-sized enterprise as defined in Commission Recommendation 2003/361/EC.[28]

The TCO is explicitly focused on terrorist content alone and does not cover other online harms.

UK Online Safety Act (UK OSA)

The UK OSA aims to protect children and adults online, imposing a "duty of care" on social media companies and services. The Act distinguishes between duties relating to 'priority illegal content' (including terrorism, child sexual exploitation and abuse, fraud, and other serious offences) and duties focused on protecting children from harmful but legal content, such as pornography or content promoting suicide or eating disorders.

Under the UK OSA, Ofcom says its codes of practice aim to take a proportionate approach, meaning the regulation imposes more stringent measures for "larger and riskier services".[29] During its passage through Parliament, an amendment to the UK OSA ensured Ofcom can consider both size or functionality when setting the thresholds for category 1 services – companies which are liable for the widest range of requirements.

Civil society organisations have raised concerns that the implementation of the regulation risks leaving loopholes for small or medium-sized but high-risk services. This may mean some services could escape categorisation that accurately reflects the level of risk they pose, regardless of the size of their userbase.[30]

Ofcom defines "small but risky" platforms as:

- Those that are "typically low reach", meaning under or around 1% of UK population as active monthly users,
- Those that "have high risk features or functionality",
- Those that are brought to Ofcom's attention for other risk factors.

In mid-2024, Ofcom created a dedicated taskforce to tackle these kinds of services. At the time of writing, information on resourcing and priority platforms was not publicly available.[31]

Australia Online Safety Act (Australian OSA)

Australia's OSA came into effect in January 2022. It significantly strengthened and extended existing laws for online safety, making online service providers more accountable for the online safety of their users. The Australian OSA defines and regulates particular types of online harms.[32]

The regulation differs to its equivalents in the UK or EU as it does not generally create tiered obligations based on platform size. Broadly, this means smaller platforms face similar baseline obligations to mainstream services, though there is acknowledgement of proportionality across types of online services providers.[33]

eSafety has schemes that allow Australians to report cyberbullying of children, adult cyber abuse, image-based abuse (sharing, or threatening to share, intimate images without the consent of the person shown), and illegal and restricted content that is class 1 or class 2.[34]

In addition, eSafety is responsible for overseeing the development of online safety industry codes for managing illegal and restricted online content.[35]

Australia's OSA also empowers eSafety to require social media services, relevant electronic services (such as messaging, gaming, and dating services), and designated internet services (other apps and websites) to report on the reasonable steps they are taking to comply with the Australian Government's Basic Online Safety Expectations (the Expectations). This is to make sure these services are transparent, accountable and safe for people to use.[36]

To ensure domestic coordination, eSafety is a member of the Digital Platform Regulators Forum (DP-REG), an information-sharing and collaboration initiative between Australian independent regulators to ensure safe, fair, innovative and competitive online spaces.

Non-governmental initiatives

Several non-governmental initiatives have made positive efforts in countering illegal content online including on smaller or emerging services. The Tech Coalition is an industry alliance of tech companies working to counter the sexual exploitation and abuse of children online, including via technical methods such as multimedia hashing and classifiers to detect and action CSAM.[37] More than 1,400 companies are also registered to make reports to the CyberTipline of the National Center for Missing and Exploited Children (NCMEC), a nonprofit based in the US that also works on CSAM and related issues.[38] The Internet Watch Foundation (IWF), a UK-based charity, maintains URL and hash lists to support the blocking of CSAM by technology companies across the industry.[39]

The Christchurch Call is an initiative that was initially established in 2019 as a commitment by governments and online service providers to eliminate terrorist and violent extremist content online. In May 2024, the initiative launched a new non-governmental organisation, the Christchurch Call Foundation (CCF).[40] The CCF includes "awareness raising and capacity building activities aimed at smaller online service providers" in its commitments. It reiterated this commitment during its 2023 Leaders' Summit, including producing a background paper on supporting smaller platforms.[41]

The Global Internet Forum to Counter Terrorism (GIFCT) is an NGO originally founded as a multistakeholder industry initiative and funded by technology companies that aims to foster collaboration among its 33 members.[42] Principal among GIFCT's offerings are its hash sharing database, used by members to detect and remove terrorist content, and its online incident response capability, such as instances of livestreamed terrorist attacks.[43] Tech Against Terrorism's Terrorist Content Analytics Platform (TCAP) detects and sends alerts to a significant number of platforms to terrorist content on their services. The tool sent a total of 10,174 alerts to 57 companies in the year covering December 2021 to December 2022, the date of its last transparency report. The report states that 82% of the content the TCAP had reported was removed following the alert. 150 platforms were registered to receive alerts from the TCAP at the time of its publication.[44]

Another new initiative with scope to counter a variety of harms, ROOST.tools (Robust Open Online Safety Tools), was launched at the Paris AI Action Summit in February 2025. The non-profit organisation has received funding from companies including Google, OpenAI, Discord and Roblox. It develops, maintains and distributes open-source online safety software such as case management systems (CMS) and review consoles. It also focuses on child safety and foundation model-powered content safeguards.[45] In 2025, the Christchurch Call Foundation announced a partnership with ROOST to adapt and focus its tool on terrorism and violent extremism-specific issues.[46] ROOST aims to support both large and small services in integrating scalable, interoperable safety structures into their platforms without having to develop such infrastructure from scratch.

There are also a range of organisations working on prevention-focused and upstream interventions, such as those focused on building user resilience or those focused more on offline criminality prevention. Examples of programmes include Inform Plus and Engage Plus from Stop it Now, a UK-based charity that works to prevent reoffending among people who have been arrested, cautioned or convicted for online offenses related to CSAM.[47] In another example, Digital Public Square is a Canada-based not-for-profit that builds tools to support communities in tackling online harms including misinformation, foreign interference and building resilience to violent extremist narratives.[48]

Gaps in the response

Most online regulatory regimes feature both proactive and reactive measures to reduce the risk of illegal or harmful content appearing on a given online service. There are systemic regulatory measures under Australia's OSA, for example, which require platforms to report on how they are meeting regulation's Basic Online Safety Expectations (BOSE).[49] According to eSafety, the requirements are designed to improve platform safety standards, as well as to improve accountability and transparency. The obligation for companies to respond to a reporting requirement is enforceable and backed by civil penalties, among other mechanisms.

Under the DSA, services that are designated as a VLOPSE must have systems in place to prevent or mitigate the risk of the dissemination of illegal content. The UK OSA, like the DSA, also imposes proactive obligations to minimise the risk of users encountering harmful or illegal content, including through safety-by design approaches, risk assessments, and measures such as algorithmic controls or access limitations. These measures are in addition to reactive requirements for platforms to remove illegal content when they are made aware of it. Under both pieces of legislation, there are tiered approaches to these duties applied on platforms, meaning the larger services are subject to more stringent measures than smaller ones.

However, tiered approaches may mean that smaller and medium-sized services are not required to take proactive steps even if they provide a safe haven for terrorists or other malicious actors (intentionally or unintentionally). Measures in the EU's DSA, for example, are guided by the scale and reach of a given service, meaning larger platforms are subject to more stringent measures than smaller ones. Only VLOPSEs (defined as having more than 45 million monthly active EU users) must conduct measures such as mandatory systemic risk assessments (Article 34), implementation of risk mitigation measures (Article 35), annual independent audits (Article 37), enhanced transparency around recommender systems (Article 27) and advertising (Article 39), the appointment of compliance officers (Article 41) and data access to vetted researchers for public interest research (Article 40).

Under the UK OSA, Ofcom says it is taking a "risk-based and proportionate" approach, although the legislation similarly applies "more onerous" requirements upon the "largest services with the highest reach and/or those services that are particularly high risk".[50] Depending on the outcome of a platform's risk assessment, they must implement a range of measures such as age controls, adult user tools and measures to mitigate the risks posed by algorithmic recommender systems.[51]

Lack of definitional clarity

However, differing perspectives on how smaller platforms should be defined can also mean the regulatory requirements placed on companies may not accurately reflect the challenges they face, or their capacity to be compliant with the law. Platforms can be categorised in diverse ways, including via metrics such as Monthly Active Users (MAU), revenue, number of staff, presence of high-risk functionalities such as livestreaming and nature of the service. However, some companies with many users may not draw significant revenue or have very few staff. Similarly some companies with comparatively small userbases may be more well-equipped to respond to abuse on their services.[52]

As the Christchurch Call Foundation has pointed out, referring to some services as "smaller" can therefore be misleading. In a 2021 paper, they referred to a definition that conceptualises this kind of service as one that "lacks the awareness and/or capacity" to respond to harmful content on its service. Notwithstanding definitional issues, expanding the range of factors and metrics considered by regulators when categorising services could enable a more proportionate, risk-based regulatory approach.

One indication of this is a challenge brought by the Wikimedia Foundation, which owns Wikipedia, in May 2025 against the lawfulness of the UK OSA's categorisations. According to the Wikimedia Foundation at the time of the announcement, categorisations were so "broadly drafted" that they would have "detrimental consequences" on the nonprofit, including negatively impacting the safety of its volunteers around the world. The platform operates a "volunteer-led model of creating and moderating content" and argues that it should not be placed into category 1 along with the largest social media platforms.[53]

The High Court of Justice in the UK dismissed Wikimedia Foundation's challenge in August, however, and Wikimedia Foundation said in an update to its blogpost in September 2025 that it would not appeal the decision.[54] Despite the ruling, the judge gave Wikipedia permission to challenge the UK OSA if it is later classified in category 1, which would impose the most stringent duties.[55] The Judgement noted that Wikipedia provides "significant value for freedom of

speech and expression" and that the imposition of regulatory duties that have a "significant impact" on its ability to operate may be illegal under UK human rights law.[56]

Tiered approaches overly focused on user numbers can therefore include larger but low risk services, while failing to effectively capture those smaller services that present significant risks to safety. Under the OSA, platform categorisation into regulatory tiers is determined by user thresholds and functionalities set through secondary legislation, which may not always reflect real-time risk. Such rigid approaches to categorising platforms risk falling behind the evolving nature of illegal harms on the internet. At the same time, approaches that uniformly apply the duties across all platforms risk unduly penalising smaller companies that do not have the same resources as the largest platforms. This may incentivise companies with fewer resources to over-remove content or to pull out of the market entirely due to fear of non-compliance.

Applicability to emerging technologies

It can also be unclear how some online regulation regimes apply to certain kinds of technologies. Most major federated or decentralised platforms lack the necessary tools for the management of illegal or harmful content at scale. The technical architecture of federated services means administrators or moderators only have access to logs that relate to their own instance and often review posts manually, meaning moderation is difficult at scale.

For some regulators, understanding how the Fediverse and other emerging technologies work can take more time than engaging the myriad other services that can more quickly be accessed, analysed and assessed. Decentralised structures can therefore present challenges to enforcement: for example, if illegal content such as a livestreamed hate crime is distributed on a specific instance, it may be unclear who should receive referrals or removal orders. Should accountability fall on the administrator of the specific instance that hosts the content, the developers who maintain the protocol or the entity responsible for maintaining the software?

Lessons may be drawn from earlier online regulatory regimes, particularly those that were amended due to shifts in the regulatory landscape. The EU's 1989 Audiovisual Media Services Directive (AVMSD) was revised in 2018 under Directive EU 2018/1808, making it applicable to video sharing platforms. Article 28b of the law requires that Member States ensure Video Sharing Platforms take appropriate steps to mitigate the risk of harm to minors by content, such as applying terms and conditions, user-reporting features, age verification systems and implementing transparent processes for the handling of user complaints.[57] The UK's equivalent of the AVMSD before Brexit, the VSP regime, was the precursor to the UK OSA. However, the number of smaller or emerging platforms in this context was comparatively low according to research published by the UK government in August 2021, which said that "only a small number of platforms" had entered the sector "and achieved scale in recent years".[58]

Coordination and communication gaps

Gaps in communication between regulators and online platforms are also likely to have contributed to a lack of awareness among some in the technology industry about the details of their regulatory obligations. An October 2024 study by the EU-funded Tech Against Terrorism Europe (TATE) initiative highlighted a lack of awareness among small and micro- Hosting Service Providers (HSPs) about the Terrorist Content Online (TCO) regulation, and several challenges those HSPs face in complying. The study found that some interviewees viewed these challenges as a disincentive to engage with the regulation – particularly if they had not yet been contacted by a competent authority.[59]

Issues with enforcement

Beyond the frameworks set out by regulatory regimes, gaps also exist in the identification and assessment of companies, and in the enforcement of regulation when companies are suspected of failing to comply. On 7 May 2025, the European Commission referred Czechia, Spain, Cyprus, Poland, and Portugal to the Court of Justice for failing to designate and/or empower a national Digital Services Coordinator (DSC) under the DSA.[60] According to European Digital Rights' (EDRi) DSC database, just three of 28 DSCs had opened investigations into particular companies at the time of writing in May 2025. Just one (the European Commission) had taken a decision based on those investigations according to the same source.[61]

Under the TCO, the EU Internet Referral Unit's transparency report for 2023 reported 239 removal orders sent by Member States via its Plateforme Européenne de Retraits des Contenus illégaux sur Internet (PERCI) platform, the majority of which appear to have been sent by the Bundeskriminalamt in Germany, based on a cross-reference of the two institutions' transparency reports.[62] PERCI is a central information system managed by Europol and used by Member States to coordinate regulatory enforcement, including referrals and removal orders, and for deconfliction purposes. The platform includes information on hosting service providers, their legal representatives, and contact points in the EU. Germany is likely to have benefitted from already having the necessary detection and alerting infrastructure under the 2018 Network Enforcement Act (Netzwerkdurchsetzungsgesetz, or NetzDG). However, many other European countries have had to set these capabilities up from scratch.

Despite limited transparency reporting by most individual Member States on their activities in this area, publicly-available information suggests the majority of Member States are yet to begin widespread enforcement of regulation by issuing removal orders.[63] Europol's 2023 transparency report instead indicates a heavy reliance on 'voluntary' referrals under the TCO, which are unenforceable and have fewer checks and balances. They instead rely on a company to review the content and take action in line with its own policies. This is likely to be because referrals are less burdensome for competent authorities than removal orders.[64]

Publicly-available data from company transparency reports also indicates inconsistencies in the frequency of removal orders received by platforms. According to X's most recent DSA transparency report in October 2024, the company received a total of just five removal orders for illegal content from four EU Member States between April and September 2024. Other mainstream platforms, including Facebook, Instagram and TikTok, reported higher numbers of removal orders, with several hundred sent to these companies respectively, according to recent transparency reports.[65] Others, including Reddit, Snap, DailyMotion and Wordpress, reported no government removal orders in their DSA transparency reports from the past 12 months.[66] The reasons for these discrepancies are unclear although they could be influenced by a range of factors, including that a significant amount of illegal content is being reported as and removed by platforms according to their own Terms of Service rather than requiring official removal orders. However, this could also reflect which platforms certain competent authorities are prioritising, the rate by which they receive reports for particular companies, an indicator of the overall volume of illegal content on those services, or even collection bias on behalf of competent authorities.

From the perspective of regulators, bilateral engagement with specific companies is time consuming and resource-intensive, an issue that is compounded when attempting to tackle illegal harms on numerous platforms. Regulators may also face security risks when engaging with companies that might be hostile towards them or their employees or that opt to publish confidential correspondence on their services. These kinds of engagements are not without their successes, however: Ofcom announced several improvements to safety measures on Bitchute in October 2023 as a result of bilateral engagement following the May 2022 Buffalo shooting under the former Video Sharing Platforms regulation. According to Ofcom, the engagement resulted in a tripling of the size of Bitchute's moderation team, an increase in moderators' working hours and a change in design of the platform to allow non-registered users to report harmful content.[67] Bitchute later went on to withdraw from the UK market in April 2025, however, citing the OSA as a reason for its decision.[68]

Balancing fairness with enforcement

Regulators must balance enforcing regulatory duties on smaller platforms without hindering innovation, competition or platform diversity among businesses and communities. Despite the apparent progress with Bitchute announced by Ofcom, the company has since become one of several platforms to cite the UK OSA as the reason for removing themselves from the UK marketplace by geo-blocking users, issuing an announcement alongside two other platforms, Gab and Kiwi Farms, in April 2025. All three platforms remain accessible to UK users with a Virtual Private Network (VPN) using an IP address located outside the UK. While there is research evidencing issues with harmful content on these platforms, several other smaller communities and forums with no such documented issues with illegal harms have taken similar action.[69] This includes communities on federated services like Mastodon, who have cited the obligations of the UK OSA as a reason for their closure or geo-blocking of UK users.[70]

The enduring 'whack-a-mole' problem

Regulators, law enforcement and other online safety practitioners continue to face the longstanding 'whack a mole' challenge, in which content, accounts or networks which are removed quickly reappear either on the same platform or migrate elsewhere. They must anticipate and respond with sufficient speed to this inevitable reappearance and migration of criminal actors on and across platforms. Improvements on a given platform invariably result in a deterioration elsewhere: while accounts or content may be deleted, the people behind them still exist. They will either repeatedly attempt to recreate accounts on the same platform or set up accounts and post their material in other spaces.

A recent example of extremist platform migrations from Telegram took place in late 2024. Changes to Telegram's policies were announced by the company in September 2024, shortly after the arrest of CEO Pavel Durov in France for offences related to the alleged proliferation of illegal content on Telegram.[71] The announcement from Durov mentioned a crackdown on illegal content and a renewed cooperation with law enforcement investigations. The implications of the announcement triggered a migration by some white supremacist extremists to SimpleX, a UK-based privacy-focused messaging app. Among the designated terrorist groups to join SimpleX since Telegram's announcement are The Base and some channels affiliated with IS.[72] At the time of writing there were no obvious signs that SimpleX had taken substantive action against these groups.

Mapping priority platforms

Given these challenges, regulators and law enforcement agencies must be able to identify, assess, prioritise and assign the correct national authorities for regulatory engagement to smaller companies that are most at risk. One way in which these actors are likely to be able to achieve this is via Europol's PERCI platform. However, an extensive non-governmental industry of organisations and private companies also exists whose work involves tracking and analysing illegal and harmful content across the internet, including on small or emerging platforms and technologies.

Closer collaboration and data-sharing between these organisations and regulators or other public sector online safety practitioners is likely to improve the speed and effectiveness of regulatory responses, particularly when it comes to identifying and prioritising platforms that are most in need of engagement. Given these efforts are often underfunded or reliant on public grants, such collaboration must include sustainable funding models.

To provide a starting point for the prioritisation and engagement of high-risk platforms via an example harm set (Islamist terrorism), ISD conducted a short mapping exercise of the top 20 platforms used by the IS group and Al-Qaeda over the past 12 months. The data is based on the total outlinks captured from core IS group and Al-Qaeda channels across platforms from April 2024 to April 2025. This list is drawn from a URL dataset (including duplicates) provided to ISD by Human Digital via its DeltaVision platform, a tool that monitors, tracks and blocks dangerous and illegal content across platforms.[73]

This mapping exercise is intended to focus particularly on jurisdictional and contact information of high-risk platforms. It aims to complement and build on other similar datasets and mapping exercises such as by ISD,[74] research from Tech Against Terrorism,[75] Dublin City University and the Tech Against Terrorism Europe (TATE) project, the European Audiovisual Observatory's MAVISE database and many others.[76]

5 of the top 20 services used by these actors were websites or platforms assessed to be likely operated by terrorist or violent extremist actors or their supporters and sympathisers. This means regulatory or law enforcement action against these services is likely to require engagement at the DNS level.[77] These websites and platforms have been left in the dataset to ensure it is an accurate representation of the use of the internet by these networks. We have redacted their names and included DNS-related information, such as the host, domain registrar, or DNS protection, rather than on the platform's parent company. The jurisdiction information for these cases relates to publicly-available WHOIS information.

Platform Name	No. URLs	Parent Company / available infrastructure information from WHOIS lookup[78]	Primary Jurisdiction
Self-hosted Islamic State Group-linked Rocketchat server	15,209	NameCheap, Inc.; Cloudflare	US
Telegram	5,807	Telegram Group Inc.	Dubai, UAE; European Digital Services Representative (EDSR), Brussels, Belgium[79]
Self-hosted Al-Qaeda-linked Rocketchat server	3,568	Cloudflare	US
Threema	2,312	Threema GmbH	Switzerland[80]
Matrix	1,762	New Vector Ltd; Element Software SARL; Element Software Inc; Element Software GmbH	France, Germany, UK, US[81]
Link shortener, assessed to be likely IS-linked	1,167	Cloudflare	San Jose, CA, US
Chirpwire[82]	930	Unknown / Registrar: Tucows Domains Inc.; DNS protection: Cloudflare	San Jose, CA, US[83]
Mediafire	890	Mediafire	Shenandoah, Texas, US
Files.fm	857	Files.fm Ltd.	Riga, Latvia[84]
Justpaste.it	674	Wise Web	Wrocław, Poland[85]
IS-linked file sharing website	622	Unknown; Cloudflare	San Jose, CA, US
Archive.org	542	Internet Archive	San Francisco, CA, US[86]
GoFile.io	533	GoFile	Paris, France[87]
Fileditchstuff.me	528	Unknown; Registrar: Tucows Domains, Inc.; DNS protection: ClouDNS.net	Tiblisi, Georgia[88]
Al-Shabaab website	440	Unknown; Registrar: Network Solutions; Name Servers: Afternic.com	San Francisco, CA, US[89]
Archive Today	317	Archive.is	Prague, Czech Republic[90]
Self-hosted IS propaganda website	316	Unknown; Registrar: Netim SAS; DNS Protection: Cloudflare	Bahamas[91]
Telegra.ph	273	Telegram Group Inc.	Dubai, UAE; European Digital Services Representative (EDSR), Brussels, Belgium[92]
Al-Shabaab website	263	Unknown; Registrar: Nicenic International Group Co., Limited	Banadir, Somalia[93]
Teleguard	247	Swisscows AG	Switzerland[94]

Recommendations

Regulators should work closely with external partners to gain access to primary data. Data-sharing will enable regulators to stay abreast of the evolving threat and prioritise efforts in a targeted and proportionate manner. Formalised collaboration and data access in which external providers are compensated will facilitate more effective and proportionate regulatory responses, informed by robust evidence.

Regulators and other government departments should consider building a list of accredited third-party datasets and technological solutions. Several such datasets and solutions exist across harm types including for CSAM (e.g. IWF, NCMEC) and terrorism (e.g. Human Digital's DeltaVision, GIFCT's hash-sharing database, Tech Against Terrorism's TCAP, Europol's PERCI). Aggregating these solutions – and the harm sets they aim to tackle – would constitute a useful resource across sectors in the online safety industry.

EU regulators and competent authorities should collaborate with researchers to build a database of platforms and their company registration details for DSA engagement and enforcement. Such a database could draw on other similar datasets, such as on Europol's PERCI platform. The dataset should be used for coordination, deconfliction and jurisdictional allocation of services between Digital Services Coordinators (DSCs). It could help to ensure that smaller, high-risk platforms meet their obligations in relation to illegal content and do not evade enforcement. It would also be helpful for identifying potential gaps in the current evidence base on the online ecosystem of platforms that play a key role in the dissemination of illegal content within the EU.

Create and utilise prioritisation risk frameworks. These risk frameworks should aim to help with the identification and prioritisation of smaller platforms and services from a regulatory perspective. These could be based on a variety of metrics including the volume of illegal content present on a service, the service's terms of service, their product offering and the rate at which they remove illegal material. Any risk matrix should be maintained in collaboration with primary data streams. These should be obtained from accredited third-party providers if they cannot be obtained in-house. The use of risk matrixes and connected prioritisation models would also help to inform regulatory triage and capacity allocation.

Anticipate side-effects of unilateral enforcement: It is crucial that bilateral engagements with specific companies are not conducted in the absence of planning for the potential side-effects, such as platform migration. Where possible, regulators and government departments should communicate with key stakeholders, including in different jurisdictions, about engagements that are likely to cause a sudden or drastic shift in the nature of the threat landscape. These communications should happen in a broader context of better information-sharing, deconfliction and coordination between regulators, law enforcement, security services and other relevant actors.

Counter illegal content at the ecosystem level: Regulatory work should aim for a model that tackles illegal content at the ecosystem level rather than only via bilateral engagement with companies. A significant body of research has shown how malicious actors – including terrorists, extremists and child abusers – use multiple platforms simultaneously. This means successful improvements on one platform are unlikely to substantively impact their operations across the other services on which they operate. Regulators and other online safety practitioners must adapt their approaches to map and intervene across networked ecosystems on multiple services.

Improve the volume and clarity of the communication of regulatory duties, particularly to smaller companies: Despite concerted efforts by Ofcom, the EU-funded TATE project and others to communicate clearly the implications of regulatory regimes to industry, there still appears to be a lack of understanding in parts of the sector. Further resources and explainers should be made publicly available or promoted, and incentives should be created for platforms to engage proactively with regulatory bodies. Awareness campaigns and outreach strategies must account for the negative perception among some technology companies of online regulation. This could potentially be achieved by shifting the narrative to better inform companies of their obligations and the benefits of early engagement with regulators to determine what steps may need to be taken to comply. Messaging must be clear and proportionate, distinguishing between supportive engagement and formal enforcement to allay concerns potentially cooperative platforms may have in taking a more proactive approach to regulatory compliance.

Conclusion

The spread of illegal or otherwise harmful content across the internet presents practitioners with an increasingly dynamic and complex challenge. This is particularly true when dealing with a growing selection of smaller or emerging platforms, many of which risk being overlooked both in online regulatory frameworks and their enforcement. Many online regimes have both proactive measures and reactive requirements to take down illegal content when it is identified and reported to platforms. However, approaches to smaller services and technologies are currently hindered by issues around prioritisation, identification, assessment and enforcement.

It is crucial that stakeholders work to apply more proactive, risk-based approaches to smaller platforms, to accurately identify, assess and enforce these requirements on high-risk platforms in a timely manner, and to do so in a way that anticipates and ideally mitigates the probable side effects on other parts of the internet. Stakeholders must continue to improve information-sharing across the sector, including related to deconfliction, approaches that have worked and lessons learned. Information-sharing should particularly apply to smaller platforms and services, including awareness-raising of their regulatory duties and the tools and resources available to assist them with compliance.

Endnotes

1 In this brief we use the term "illegal hate speech" to align with the EU Digital Services Act (DSA), where it constitutes a category of illegal content. The DSA itself does not create a new definition, but refers to existing EU and Member State laws, most notably the 2008 Framework Decision on combating racism and xenophobia. This should be distinguished from other jurisdictions: in the UK, hate offences fall under the category of "priority illegal content" in the Online Safety Act; in Australia, relevant provisions are spread across online safety and anti-discrimination legislation; while in other jurisdictions, hate speech can be protected by constitutions and only be unlawful where it falls into narrower categories such as incitement or true threats..

2 eSafety Commissioner. The Global Online Safety Regulators Network. Retrieved from: https://www.esafety.gov.au/about-us/consultation-cooperation/international-engagement/the-global-online-safety-regulators-network.

3 Ofcom. (24 January 2025). Helping small services navigate the Online Safety Act". Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/helping-small-services-navigate-the-online-safety-act. .

4 See, for example: Macdonald, S & McCafferty, S. (2024). Online Jihadist Propaganda Dissemination Strategies. Retrieved from: https://voxpol.eu/wp-content/uploads/2024/03/DCU-PN0752-Online-Jihadist-WEB-240305.pdf; Internet Watch Foundation. (2018). Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed. Child Sexual Abuse. Retrieved from: https://www.iwf.org.uk/media/23jj3nc2/distribution-of-captures-of-live-streamed-child-sexual-abuse-final.pdf; Mulhall, J. (17 March 2022). I'll be back: The Rise of Far-Right Alt-Tech. Retrieved from: https://hopenothate.org.uk/2022/03/17/ill-be-back-the-rise-of-far-right-alt-tech/.

5 Fischer, A. & Prucha, N. (2022). The Salafi-Jihadi online ecosystem in 2022. Retrieved from: https://eictp.eu/the-salafi-jihadi-online-ecosystem-2022-swarmcast-2-0/.

6 Ofcom. (3 October 2023). BitChute: compliance assurances to protect users from videos containing harmful material. Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/bitchute-03-10-2023.

7 Department for Science, Innovation, and Technology, UK government. (8 May 2025). Draft Statement of Strategic Priorities for Online Safety. Retrieved from: https://www.gov.uk/government/publications/statement-of-strategic-priorities-for-online-safety/statement-of-strategic-priorities-for-online-safety.

8 Ofcom. (24 April 2025). Overview: Protecting people from illegal harms online. Retrieved from: https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/overview.pdf?v=387529.

9 Ofcom. (7 March 2025). Enforcement programme into measures being taken by file-sharing and file-storage services to prevent users from encountering or sharing child sexual abuse material (CSAM). Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/enforcement-programme-into-measures-being-taken-by-file-sharing-and-file-storage-services-to-prevent-users-from-encountering-or-sharing-child-sexual-abuse-material-csam; National Center for Missing & Exploited Children. Child Sexual Abuse Material. Retrieved from: https://www.missingkids.org/theissues/csam.

10 ISD. (2023). Emerging Platforms and Technologies: An Overview of the Current Threat Landscape and its Policy Implications. Retrieved from: https://www.isdglobal.org/isd-publications/emerging-platforms-and-technologies-an-overview-of-the-current-threat-landscape-and-its-policy-implications/.

11 Robertson, A. (12 July 2019). How the biggest decentralised social network is dealing with its Nazi problem. The Verge. Retrieved from: https://www.theverge.com/2019/7/12/20691957/mastodon-decentralized-social-network-gab-migration-fediverse-app-blocking; Leidig, E. (17 February 2021). Odysee: The New YouTube for the Far-right. GNET. Retrieved from: https://gnet-research.org/2021/02/17/odysee-the-new-youtube-for-the-far-right/; Schlegel, L. et al. (12 February 2025). Exploring the digital extremist ecosystem: a preliminary analysis of hateful posts on Mod DB. Frontiers in Psychology, Vol. 15. Retrieved from: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1502098/full; Thompson, E. and Lamphere-Englund, G. (November 2024). 30 years of trends in terrorist and extremist games. GNET. Retrieved from: https://gnet-research.org/wp-content/uploads/2024/10/GNET-47-Extremist-Games_web.pdf. Anti-Defamation League. (21 April 2025). The Dark Side of Roblox: 'Active Shooter Studios' Create Maps Based on Real-Life Mass Shootings. Retrieved from: https://www.adl.org/resources/article/dark-side-roblox-active-shooter-studios-create-maps-based-real-life-mass.

12 Amarasingam, A. et al. (July 2022). The Buffalo attack: The cumulative momentum of far-right terror. CTC Sentinel. Retrieved from: https://ctc.westpoint.edu/wp-content/uploads/2022/07/CTC-SENTINEL-072022.pdf.

13 Grayson, N. (20 May 2022). How Twitch took down Buffalo shooter's stream in under two minutes. The Washington Post. Retrieved from: https://www.washingtonpost.com/video-games/2022/05/20/twitch-buffalo-shooter-facebook-nypd-interview/.

14 Global Internet Forum to Counter Terrorism. Membership. Retrieved from: https://gifct.org/membership/.

15 Fitzpatrick, A. (21 January 2021). Why Amazon's Move to Drop Parler is a big deal for the future of the internet. Time. Retrieved from: https://time.com/5929888/amazon-parler-aws/.

16 Townsend, M. (17 January 2021). How Trump supporters are radicalised by the far-right. Retrieved from: https://www.theguardian.com/technology/2021/jan/17/how-trump-supporters-are-radicalised-by-the-far-right.

17 Fiennes, G. (27 March 2025). The Great TikTok Migration: Western Extremists Flock to RedNote. Retrieved from: https://gnet-research.org/2025/03/27/the-great-tiktok-migration-western-extremists-flock-to-rednote/´.

18 Baptista, E., Hu, K. and Oladipo, D. (15 January 2025). Over half a million 'TikTok refugees' flock to China's RedNote. Retrieved from: https://www.reuters.com/technology/over-half-million-tiktok-refugees-flock-chinas-rednote-2025-01-14/; Fiennes, G. (27 March 2025). The Great TikTok Migration: Western extremists flock to RedNote. Retrieved from: https://gnet-research.org/2025/03/27/the-great-tiktok-migration-western-extremists-flock-to-rednote/.
19 European Commission. The Digital Services Act. Retrieved from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.
20 Ibid.
21 Tremau. A Guide for All Online Services: DSA Whitepaper. Retrieved from: https://admin.tremau.com/wp-content/uploads/2025/03/DSA-Whitepaper-final-2.pdf.
22 Tremau. (7 December 2023). What does the DSA mean for your business? Retrieved from: https://tremau.com/resources/what-does-the-dsa-mean-for-your-business/.
23 Latham & Watkins. (March 2023). The Digital Services Act: Practical Implications for Online Services and Platforms. Retrieved from: https://www.lw.com/admin/upload/SiteAttachments/Digital-Services-Act-Practical-Implications-for-Online-Services-and-Platforms.pdf.
24 Tech Against Terrorism. Terrorist Content Online (TCO) Guide. Retrieved from: https://www.techagainstterrorism.org/hubfs/TCO-Guide.pdf.
25 O'Carroll, L. (13 November 2024). Ireland orders X, TikTok and Instagram to curb terrorist content. Retrieved from: https://www.theguardian.com/world/2024/nov/13/ireland-orders-x-tiktok-and-instagram-to-curb-terrorist-content.
26 European Commission, Directorate-General for Migration and Home Affairs. (April 2021). Terrorist Content Online Regulation (Regulation (EU) 2021/784): Questions & Answers. Retrieved from: https://home-affairs.ec.europa.eu/system/files/2021-05/202104_terrorist-content-online_en.pdf
27 O'Carroll, L. (13 November 2024). Ireland orders X, TikTok and Instagram to curb terrorist content. Retrieved from: https://www.theguardian.com/world/2024/nov/13/ireland-orders-x-tiktok-and-instagram-to-curb-terrorist-content.
28 European Commission. Commission Recommendation 2003/361/EC of 6 May 2003: Definition of Micro, Small and Medium-sized Enterprises (OJ L 124, pp. 36–41). Retrieved from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:124:0036:0041:en:PDF.
29 Ofcom. (30 January 2024). Why size and risk matter in our approach to online safety. Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/why-size-and-risk-matter-in-our-approach-to-online-safety.
30 Woods, L. and Walsh, M. (1 May 2024). Categorisation of services in the Online Safety Act. Retrieved from: https://www.onlinesafetyact.net/analysis/categorisation-of-services-in-the-online-safety-act/; Antoniou, A. (19 October 2023). Bringing small high-harm platforms into the online safety regime: how one word changed the game. Retrieved from: https://www.onlinesafetyact.net/analysis/bringing-small-high-harm-platforms-into-the-online-safety-regime-how-one-word-changed-the-game/.
31 Ofcom. (28 April 2025). Freedom of Information Request. Retrieved from: https://www.ofcom.org.uk/siteassets/resources/documents/about-ofcom/foi/2025/april/online-safety-small-but-risky.pdf?v=396599.
32 eSafety can investigate complaints from individuals and help to stop, remove and limit the impacts of cyberbullying of children, adult cyber abuse, image-based abuse (sharing or threatening to share intimate images and videos without the consent of the person shown), as well as illegal and restricted online content. See here: https://www.esafety.gov.au/about-us/industry-regulation and here: https://www.esafety.gov.au/report.
33 eSafety Commissioner. (February 2022). Online Safety Act 2021 Fact Sheet. Retrieved from: https://www.esafety.gov.au/sites/default/files/2022-02/OSA%20fact%20sheet%20updated.pdf?v=1747903514413.
34 eSafety Commissioner. Report Online Harm. Retrieved from: https://www.esafety.gov.au/report.
35 Phase 1 Industry Codes and Standards regulate Class 1A and 1B material, covering the most seriously harmful online content, including material that shows sexual abuse of children and acts of terrorism. They contain enforceable obligations that apply to eight sections of the online industry. Phase 2 industry codes have been under development since mid-2024, and as of June 2025, eSafety has registered codes for search engines, hosting services and internet carriage services; the remaining codes are still under regulatory assessment. See: eSafety Commissioner. Industry Codes and Standards (last updated 12 September 2025). Retrieved from: https://www.esafety.gov.au/industry/codes.
36 eSafety Commissioner. Basic Online Safety Expectations. Retrieved from: https://www.esafety.gov.au/industry/basic-online-safety-expectations.
37 Technology Coalition. (2024). Retrieved from: https://www.technologycoalition.org/newsroom/2024-reports.
38 National Center for Missing & Exploited Children. Child Sexual Abuse Material (CSAM). Retrieved from: https://www.missingkids.org/theissues/csam.
39 Internet Watch Foundation. URL List. Retrieved from: https://www.iwf.org.uk/our-technology/our-services/url-list/; Image Hash List. Retrieved from: https://www.iwf.org.uk/our-technology/our-services/image-hash-list/.
40 French Presidency / Élysée. (14 May 2024). Joint press release between France and New-Zealand — Launch of the Christchurch Call Foundation. Retrieved from: https://www.elysee.fr/en/emmanuel-macron/2024/05/14/joint-press-release-between-france-and-new-zealand-launch-of-the-christchurch-call-foundation.
41 Christchurch Call. Leaders Summit 2023 — Background paper: Supporting smaller platforms. Retrieved from: https://www.christchurchcall.org/content/files/2024/06/Leaders-Summit-2023-Background-paper-Supporting-smaller-platforms.pdf.
42 GIFCT. (2024). Annual Report 2023. Retrieved from: https://gifct.org/wp-content/uploads/2024/04/GIFCT-Annual-Report-2023.pdf.
43 GIFCT. Homepage / About. Retrieved from: https://gifct.org/.

44 "Terrorist Content Analytics Platform. Transparency Report December 2021 - November 2022", Retrieved from: https://terrorismanalytics.org/policies/transparency-report.

45 Roost. Modular safety solutions for builders. Retrieved from: https://roost.tools/; Newton, C. (10 February 2025). An overdue idea for making the internet safer just got the funding it needs. Retrieved from: https://www.platformer.news/roost-open-source-trust-safety/.

46 Christchurch Call. Christchurch Call partners with Roost for open-source AI tools to tackle TVEC. Retrieved from: https://www.christchurchcall.org/christchurch-call-partners-with-roost-for-open-source-ai-tools-to-tackle-tvec/.

47 Stop It Now / Lucy Faithfull Foundation. Inform Plus and Engage Plus — for people who have offended online. Retrieved from: https://www.stopitnow.org.uk/inform-plus-and-engage-plus-for-people-who-have-offended-online/.

48 Digital Public Square. Our Work. Retrieved from: https://digitalpublicsquare.org/our-work/.

49 eSafety Commissioner. Basic Online Safety Expectations. Retrieved from: https://www.esafety.gov.au/industry/basic-online-safety-expectations.

50 Ofcom. (24 January 2025). Helping small services navigate the online safety act. Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/helping-small-services-navigate-the-online-safety-act.

51 Department for Science, Innovation and Technology. (24 April 2025). Online Safety Act: Explainer: Department for Science, Innovation, and Technology. Retrieved from: https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer.

52 Christchurch Call. (June 2024). Supporting Smaller Platforms. Retrieved from: https://www.christchurchcall.org/content/files/2024/06/Leaders-Summit-2023-Background-paper-Supporting-smaller-platforms.pdf.

53 Wikimedia Foundation. (19 September 2023). Wikimedia Foundation calls for protection and fair treatment of Wikipedia as UK Online Safety Bill becomes law. Retrieved from: https://wikimediafoundation.org/news/2023/09/19/wikimedia-foundation-calls-for-protection-and-fair-treatment-of-wikipedia/.

54 Wikimedia Foundation. (8 May 2025). Wikimedia Foundation brings legal challenge to new UK Online Safety Act requirements. Retrieved from: https://diff.wikimedia.org/2025/05/08/wikimedia-foundation-brings-legal-challenge-to-new-uk-online-safety-act-requirements/#.

55 Wikimedia Foundation. (12 September 2025). Wikimedia Foundation challenges UK Online Safety Act regulations. Retrieved from: https://wikimediafoundation.org/news/2025/09/12/wikimedia-foundation-challenges-uk-online-safety-act-regulations/.

56 Judiciary of England & Wales. Wikimedia Foundation and another v Secretary of State for Science, Innovation and Technology (Approved Judgment, 11 August 2025) [2025] EWHC 2086 (Admin). Retrieved from: https://www.judiciary.uk/wp-content/uploads/2025/08/Wikimedia-Foundation-and-another-v-Secretary-of-State-for-Science-Innovation-and-Technology.pdf.

57 European Parliament. (October 2022). Transposition of the 2018 Audiovisual Media Services Directive. Retrieved from: https://www.europarl.europa.eu/RegData/etudes/IDAN/2022/730354/EPRS_IDA(2022)730354_EN.pdf.

58 Department for Science, Innovation and Technology & Department from Digital, Culture, Media & Sport. (3 August 2021). Summary factsheet: Understanding how platforms with video-sharing capabilities protect users from harmful content online. Retrieved from: https://www.gov.uk/government/publications/understanding-how-platforms-with-video-sharing-capabilities-protect-users-from-harmful-content-online/summary-factsheet-understanding-how-platforms-with-video-sharing-capabilities-protect-users-from-harmful-content-online.

59 Tech Against Terrorism Europe. (8 October 2024). The challenges that small and micro HSPs face in implementing the TCO regulation. Retrieved from: https://tate.techagainstterrorism.org/news/report-the-challenges-that-small-and-micro-hsps-face-in-implementing-the-tco-regulation.

60 European Commission. (7 May 2025). Commission decides to refer Czechia, Spain, Cyprus, Poland and Portugal to the Court of Justice of the European Union due to lack of effective implementation of the Digital Services Act.Retrieved from: https://digital-strategy.ec.europa.eu/en/news/commission-decides-refer-czechia-spain-cyprus-poland-and-portugal-court-justice-european-union-due.

61 The Digital Services Coordinators Database. Retrieved from: https://dscdb.edri.org/.

62 Europol. (Update date: 9 January 2025). EU Internet Referral Unit Transparency Report 2023. Retrieved from: https://www.europol.europa.eu/publications-events/publications/eu-internet-referral-unit-transparency-report-2023; Bundeskriminalamt. Transparenzbericht für das Jahr 2023 zur Umsetzung der TCO-Verordnung. Retrieved from: https://www.bka.de/DE/UnsereAufgaben/Deliktsbereiche/PMK/TCO-VO/TCO-VO_node.html.

63 EDRi. (11 September 2024). Denmark will issue removal orders without court approval: Impacts on free speech and pro-Palestinian voices. Retrieved from: https://edri.org/our-work/denmark-will-issue-removal-orders-without-court-approval-impacts-on-free-speech-and-pro-palestinian-voices/.

64 Europol. (Update date: 9 January 2025). EU Internet Referral Unit Transparency Report 2023. Retrieved from: https://www.europol.europa.eu/publications-events/publications/eu-internet-referral-unit-transparency-report-2023.

65 Europol. (2025). 2023 EU Internet Referral Unit Transparency Report. Retrieved from:: https://www.europol.europa.eu/cms/sites/default/files/documents/2023%20EU%20Internet%20Referral%20Unit%20Transparency%20Report.pdf and Meta. Regulatory Transparency Reports. Retrieved from: https://transparency.meta.com/reports/regulatory-transparency-reports/.

66 X (formerly Twitter). (covers period 1 April to 30 September 2024). DSA Transparency Report – October 2024. Retrieved from: https://transparency.x.com/dsa-transparency-report.html and Reddit. EU Terrorist Content Online Regulation Transparency Report 2024. Retrieved from: https://40687240.fs1.hubspotusercontent-na1.net/hubfs/40687240/Reddit%E2%80%99s%20EU%20Terrorist%20Content%20Online%20Regulation%20Transparency%20Report%20(2024).pdf and Squarespace / Platform (publisher unspecified). (28 February 2025). Digital Services Act Report. Retrieved from: https://static1.squarespace.com/static/5134cbefe4b0c6fb04df8065/t/67c72707c487fd-

31168cbe5d/1741104903826/Digital+Services+Act+Report+2025.02.28.pdf and OECD. (June 2024). Transparency Reporting on Terrorist and Violent Extremist Content Online, Fourth Edition. Retrieved from: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/transparency-reporting-on-terrorist-and-violent-extremist-content-online_3f72a170/901cb8cf-en.pdf.

67 Ofcom. (3 October 2023). BitChute improves its safety measures following engagement with Ofcom. Retrieved from: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/bitchute-improves-its-safety-measures-following-engagement-with-ofcom.

68 Bitchute. UK Regulation. Retrieved from: https://www.bitchute.com/ukregulation.

69 Hern, A. (7 September 2022). TechScape: How Kiwi Farms, the worst place on the web, was shut down. Retrieved from: https://www.theguardian.com/technology/2022/sep/07/techscape-kiwi-farms-cloudflare; Davis, G. (20 July 2020). Bitchute: Platforming Hate and Terror in the UK. Retrieved from: https://hopenothate.org.uk/2020/07/20/bitchute-platforming-hate-and-terror-in-the-uk/; Fighting Online Antisemitism. (17 March 2025). Antisemitism on Gab. Retrieved from: https://www.worldjewishcongress.org/en/news/gabs-unmoderated-platform-fuel-antisemitic-hate-warns-foa-and-wjc-report.

70 Online Safety Act. In memoriam. Retrieved from: https://onlinesafetyact.co.uk/in_memoriam/; Masnick, M. (20 December 2024). Death of a Forum: How the UK's Online Safety Act is killing communities. Retrieved from: https://www.techdirt.com/2024/12/20/death-of-a-forum-how-the-uks-online-safety-act-is-killing-communities/.

71 Collier, K. and Wile, R. (28 August 2024). Telegram CEO charged by French prosecutors. Retrieved from: https://www.nbcnews.com/tech/tech-news/telegram-ceo-pavel-durov-charged-french-prosecutors-rcna168603.

72 Makuch, B. (4 October 2024). Far-right extremists flee from Telegram to SimpleX over privacy features. Retrieved from: https://www.theguardian.com/technology/2024/oct/04/telegram-simplex-far-right.

73 DeltaVision. Solutions for Harmful Online Content. Retrieved from: https://deltavision.io/.

74 Institute for Strategic Dialogue. (2022). Researching the Evolving Online Ecosystem: Barriers, Methods and Future Challenges. Retrieved from: https://www.isdglobal.org/wp-content/uploads/2022/07/Researching-the-Evolving-Online-Ecosystem_Main-report.pdf ISD and Institute for Strategic Dialogue. (2022). Researching the Evolving Online Ecosystem: Annex. Retrieved from: https://www.isdglobal.org/wp-content/uploads/2022/07/Researching-the-Evolving-Online-Ecosystem_Annex.pdf.

75 Tech Against Terrorism. (2024). Mapping Far-Right Terrorist Propaganda Online. Retrieved from: https://techagainstterrorism.org/hubfs/TCAP_Report_Mapping_Far-right_Terrorist_Propaganda_Online%20(1).pdf.

76 European Audiovisual Observatory. MAVISE Database. Retrieved from: https://mavise.obs.coe.int/.

77 Europol removes extremist content", , 22 May 2022, Retrieved from: https://en.europarabct.com/?p=78882.

78 European Center for Counterterrorism and Intelligence Studies. Publicly-available WHOIS information is provided here in instances where the identity and location of the parent company could not be identified, and/or in cases where the platform or service is believed to be terrorist operated or linked. In these latter cases, the location included in the "Jurisdiction" column reflects the information on the WHOIS lookup related to the website's infrastructure providers, rather than the location of the registrant.

79 Haeck, P. (6 May 2024). Belgium to supervise Telegram under EU content law. Retrieved from: https://www.politico.eu/article/belgium-to-supervise-telegram-under-eus-content-law/.

80 Threema. Contact / About. Retrieved from: https://threema.com/en/contact.

81 Element (Element.io). Privacy Policy. Retrieved from: https://element.io/privacy.

82 Likely to be Al-Qaeda-linked.

83 DomainTools. Whois – Chirpwire.net. Retrieved from: https://whois.domaintools.com/chirpwire.net.

84 Files.fm. Contacts / About. Retrieved from: https://files.fm/contacts.

85 JustPaste.it. Privacy Policy. Retrieved from: https://justpaste.it/privacypolicy.

86 Internet Archive. Terms of Use (last updated December 2014). Retrieved from: https://archive.org/about/terms.php.

87 Whois.com. Whois – Gofile.io. Retrieved from: https://www.whois.com/whois/gofile.io and RIPE NCC / DB Web UI. Organization Lookup. Retrieved from: https://apps.db.ripe.net/db-web-ui/lookup?source=ripe&key=ORG-GA1190-RIPE&type=organisation.

88 Whois.com. Whois – fileditchstuff.me. Retrieved from: https://www.whois.com/whois/fileditchstuff.me.

89 DomainTools. Whois – shahadaagency.com. Retrieved from: https://whois.domaintools.com/shahadaagency.com.

90 Location based on publicly-available DNS information. The accuracy of this information could not be verified: DomainTools. Whois – archive.is. Retrieved from: https://whois.domaintools.com/archive.is

91 Whois – obedientsupporters.co. Retrieved from: https://www.whois.com/whois/obedientsupporters.co.

92 Telegram. (22 November 2016). Instant View, Telegraph, and Other Goodies. Retrieved from: https://telegram.org/blog/instant-view

93 Whois – shahadanews.info. Retrieved from: https://www.whois.com/whois/shahadanews.info.

94 Swisscows AG. Swisscows – Company / About. Retrieved from: https://company.swisscows.com/en.

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2025). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address 3rd Floor, 45 Albemarle Street, Mayfair, London, W1S 4JL. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved.

www.isdglobal.org