Organization for Security and Co-operation in Europe
The Representative on Freedom of the Media

Online Safety Bill (as per HL Bill 170 Commons Amendments to Lords Amendments, Consequential Amendments, Disagreements, Amendments in Lieu and Reasons, 13 September 2023)

Legal Review

Commissioned by the Office of the OSCE Representative on Freedom of the Media from
Dr Paolo Cavaliere, Senior Lecturer in Digital Media and IT Law
University of Edinburgh Law School

October 2023

Table of Contents

I. Executive summary

Scope and jurisdiction

The Online Safety Act, passed in September 2023, is a comprehensive legal framework that sets up a wide range of obligations for digital platforms, search engines and similar service providers, including among others monitoring, reporting and removal obligations. In the Government's words, the Act pursues a 'zero-tolerance approach' against content that could prove harmful to children and help adults to 'take control of their online lives, while protecting our mental health.'[1] Regulated service providers will face 'significant fines that could reach billions of pounds [and i]n some cases, their bosses may even face prison.'[2]

The Act foresees a graduated approach, dividing service providers into three categories depending on their size, functionalities offered, and the level of risk they pose. The Act also aims to have extra-territorial application as it deems for application to service providers with 'links' in the UK.

Concerns for media freedom

Throughout the Act, several provisions are covered by what appears as a 'news exemption' which aims to exclude the applicability of such norms to content originating from news media outlets. The definition of 'recognised news publisher' is evidently focused on institutional media outlets and is likely to run against both the general principles of a graduated and differentiated approach towards media regulation as recommended by the Council of Europe and the functional approach to defining journalism emerging from the case-law of the European Court of Human Rights and, as such, risks to leave out a series of outlets and other media actors that play an important role in providing information of public interest to the public and giving voice to more marginalised communities. However, the Act also includes a duty for service providers to protect content of democratic importance and the Introduction includes a helpful reference to 'users' rights to freedom of expression and privacy'.

Regulator's independence

The independent industry regulator OFCOM will be tasked with drafting and amending codes of conduct, while the Secretary of State and the Parliament will nonetheless have extensive powers to determine the content of such codes. OFCOM does not appear to be adequately safeguarded from government interfering with its own regulatory powers and does not seem to be granted a significant margin to exercise its own discretion.

Monitoring duties

The Act establishes several duties for service providers in respect of different kinds of content. Overall, while the breadth and scope of such monitoring duties seems necessary to achieve the ambitious goals of this regulatory effort, the provisions seem to lack the necessary degree of precision to allow the regulated entities to understand the extent of their obligations and the concrete risk of incurring into penalties. Some of the respective provisions are formulated in quite generic language (a duty to take down illegal content 'swiftly', an expectation that service providers make use of 'design …, algorithms and other features'). The timing and the methods

[1] Press release, Britain makes internet safer, as Online Safety Bill finished and ready to become law - GOV.UK (www.gov.uk) , 19 September 2023.
[2] Ibid.

that service providers will need to recur to will need to be specified in greater detail or else uncertainty over the risk of penalties might offer an incentive to over-removals.
OFCOM is also granted the power to impose economic contraventions and penalties in different forms, and can also request regulated services to 'take steps' to comply with a notified requirement or remedy a failure to comply with it, including the deployment of any 'proactive technology'. This provision appears particularly broad and vague, most immediately in regard to the type of technology that service providers would be expected to use; furthermore, it appears that such an obligation could easily turn into a proactive monitoring obligation. It is concerning that, by including in legal frameworks generic references to unspecified technologies, new and increasingly more active and general forms of monitoring become progressively normalised, silently eroding existing safeguards for freedom of expression and the free flow of information.

Content removal obligations

The Act identifies certain specific categories of content and corresponding actions expected from service providers. Overall, the Act foresees a broad and wide-ranging list of types of content on which service providers are expected to intervene. In many cases, there appears to be a basis in existing statutory provisions, while the Act also creates a few new offences (such as the false communications offence, the threatening communications offence, sending or showing flashing images, and criminalising assisting or encouraging self-harm online). It remains key, however, that any such provisions are interpreted as narrowly as possible, according to an equally compulsory principle of proportionality repeatedly and consistently affirmed by the European Court of Human Rights in its case-law. This especially concerns those wordings which could more easily lend themselves to overbroad and expansive interpretation (e.g. 'public order offences' or 'assisting illegal immigration').
The category of 'false communication offences' appears problematic. In fact, the provision fails to determine what specific harm a communication should be capable to cause in order to fall under this provision; and why, if harmful anyway, such communications would also need to be 'false' in order to be restricted. In view of these questions, the possibility of imposing prison sentences for such offences is likely to be incompatible with international standards and recommendations.

Service providers' own terms of service and reporting obligations

Service providers can take down content, restrict users' access to information, or suspend or ban users using 'proportionate systems and processes' and 'in accordance with the terms of service' which should be 'clear and accessible'. OFCOM is expected to produce the respective guidance for providers.
OFCOM in this sense will play a fundamental role as the Act provides, at present, little guidance on the circumstances under which platforms could deploy their own terms of service as the basis for removal decisions; instead, they will need more guidance on what instances of content moderation and/or removal service providers are requested or allowed to take in pursuit of public policy goals, and which ones they are allowed to take in pursuit of different objectives, including those related to their own business models.

Age verification and end-to-end encryption

Service providers will be required 'to use age verification or age estimation (or both) to prevent children of any age from encountering primary priority content that is harmful to children which the provider identifies on the service.' However, the requirement for service providers to verify users' age raises concerns for its foreseeable impact on the right to privacy and data protection. As a result, service providers will have to either utilise ineffective technologies, or resort to systems that would breach their users' fundamental rights (the right to privacy most immediately). The concern is that, for lack of viable and trustworthy technological means, service providers could simply decide to rid their platforms of content indicated in Sec. 61 and 62 for all their users – including adults.

The Act also allows Ofcom to require service providers to use 'accredited technology' to detect, identify and prevent individuals from accessing terrorism and CSEA content. The wording of this provision could be interpreted as allowing OFCOM to require service providers to deploy detection tools capable of circumventing human end-to-end encryption. Recently, the Secretary of State for Science, Innovation and Technology admitted that technology enabling such controls without unduly impinging on users' privacy is still in development. It thus remains important that OFCOM refrains from mandating the use of over-intrusive technologies that could have a destructive impact on fundamental rights (the right to privacy most immediately).

II. Specific recommendations

- The Secretary of State and OFCOM, while periodically revising the thresholds of each category of service providers, sould consider providing a timeline for periodic revisions, so as to guarantee that regulatory requirements will remain unchanged at least for a set period and changes would happen at a foreseeable time. OFCOM could offer support and advice, when needed, to services in the transition from one category to another.
- OFCOM should provide clear indications of the circumstances under which a foreign service provider would be deemed to have sufficient links to the UK to fall under the scope of the respective legal provisions.
- Service providers and any authorities implementing the Act must interpret it in line with the principles established by the Council of Europe and the European Court of Human Rights, including by interpreting the notions of news publishers and journalistic content from a perspective that takes into account the role and relevance that content put forward by non-professional journalistic endeavours can have for the public debate and society at large, rather than by looking solely at the characteristics of the institutions publishing the content at stake, to the extent that is possible for the purposes of this legislation. Guidance already provided by the Information Commissioner's Office in the context of the Data Protection Act would prove most helpful in this case too.[3]
- It would be recommendable to consider ways to provide for enhanced protection of the regulator's independence from the executive power.
- OFCOM could provide more specific and detailed guidance on the nature of the specific monitoring duties imposed on the service providers, in particular regarding the nature of the technology that such providers would be expected to deploy, and the exact timing of such obligations. In providing such guidance, it would be ideal if OFCOM could operate in collaboration with industry players to ascertain the feasibility of different options and with a particular view of avoiding excessive regulatory burdens.
- OFCOM (or judicial authorities) should provide guidance to service providers on what technologies to deploy, and on what 'specific elements' such technologies would need to be deployed to look for. OFCOM could collaborate routinely with service providers to identify the technologies that service providers would be expected to deploy, and the specific kinds of usage that would not unduly impinge on fundamental rights, so as to keep the guidance up to date with technological development.
- More guidance will be needed to indicate in what instances service providers will be allowed to remove content in pursuit of public policy goals, and in which ones they are expected to take action in pursuit of different objectives, including those related to their own business models.
- Service providers should be expected to clearly indicate when content is removed in compliance with legal obligations and when measures are taken in compliance with their own terms of service.
- OFCOM should provide detailed and specific guidance as to the information that service providers will be required to provide in compliance with their transparency obligations. Such guidance should provide, in particular, a common framework for all regulated

[3] Information Commissioner's Office, Draft data protection and journalism code of practice, 2022, , especially pp. 11-16.

services to follow, and should include specific requirements regarding the format in which data needs to be provided, the level of granularity of information in order to ensure a sufficient degree of uniformity across the different types of providers.

- OFCOM should develop guidance in collaboration with the Information Commissioner's Office in relation to the use of age verification systems, to minimise the amount of data collected and make the verification system reliable. Considering an independent trusted third-party system would be the most welcome option. OFCOM could also consider developing guidance on how service providers could better tailor their actions in respect to different age groups.
- OFCOM should refrain from mandating the use of technologies that, by allowing detection tools capable of circumventing human end-to-end encryption, could destructively impact on fundamental rights (right to privacy). Collaborating with industry stakeholders and developing a regulatory framework to ensure that any technology uptake in the future is placed within a framework that is fully respectful of privacy and data protection rights should remain a priority aim.

III. International legal standards concerning freedom of expression, freedom of information and other relevant principles of online content governance

International legal standards on freedom of expression

Freedom of expression and of the media are protected by relevant provisions in fundamental rights treaties at both the regional and international level. Article 19 of the International Covenant on Civil and Political Rights and Article 10 of the European Convention of Human Rights closely mirror each other as they both conceive the right to freedom of expression as comprising three distinct layers, and provide for a similar framework to impose restrictions.

In regard to the structure of the right, both the provisions break down the right into three distinct layers – an unfettered freedom to hold opinions and two qualified rights to receive (a passive aspect) and to impart (an active aspect) information and ideas. The first paragraph of Art 10 ECHR guarantees these latter two layers of the right against interferences by public authorities and regardless of frontiers; Art 19 ICCPR also established a principle of technological neutrality in that it clarifies that the exercise of the right does not depend on the specific medium utilised.

In regard to limitations, the second paragraph of Art 10 ECHR introduces the notion that 'formalities, conditions, restrictions or penalties' may be imposed under certain limited circumstances, and provides for a general framework to assess the compatibility of any such restrictions with the Convention. The framework consists of a three-part test requiring any restriction to have a basis in the law of the country where it is imposed (test of legality), to be in pursuit of one of the aims listed exhaustively in paragraph 2, i.e. 'national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary' (test of necessity) and to be the least intrusive measure possible among those effective enough to reach the designated objective (test of proportionality). Paragraph 3 of Art 19 ICCPR contains a shorter list of legitimate aims (the respect of the rights or reputations of others; the protection of national security or of public order (ordre public), or of public health or morals), however restrictions must undergo similar tests of necessity and proportionality.[4] The provisions from Art 10 ECHR and Art 19 ICCPR mirror very closely those of other regional treatises for fundamental rights, including Article 11 of the EU Charter of Fundamental Rights.

The OSCE Decision on the Safety of Journalists makes express references to the provisions in Art 19 ICCPR and its framework for restrictions to freedom of expression, in particular the circumstance that 'any restrictions on the right to freedom of expression may only be such as are provided by law and are necessary on the grounds set out in paragraph 3 of Article 19 of the ICCPR', and calls on participant States to '[f]ully implement … their international obligations related to freedom of expression and media freedom, including by respecting, promoting and protecting the freedom to seek, receive and impart information regardless of frontiers [and to b]ring their laws, policies and practices, pertaining to media freedom, fully in compliance with their international obligations and commitments.'[5]

[4] Human Rights Committee, 'General comment No. 34 to Article 19: Freedoms of opinion and expression', CCPR/C/GC/34, 2011, paras. 33-34.
[5] OSCE, Decision no. 3/18 on the Safety of Journalists, MC.DEC/3/18, 2018, p. 3.

With specific reference to the more recent issues of disinformation and so-called 'fake news', the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has expressed concern for the use of 'vague and overly broad laws to criminalize, block, censor and chill online speech'.[6] Therefore, '[c]riminal law should be used only in very exceptional and most egregious circumstances of incitement to violence, hatred or discrimination';[7] even under the most distressing circumstances such as armed conflict (and logically even more so at times of peace or in the absence of comparable major disruptions to democratic life), State authorities 'should not prohibit or restrict disinformation, propaganda and "false news" or "fake news" unless they meet the requirements of legality, necessity and legitimate aim as set out in article 19 (3) or amount to incitement in line with article 20 of the International Covenant on Civil and Political Rights'.[8] The UN Rapporteur advises that '[t]ackling disinformation requires multidimensional, multi-stakeholder responses that are well grounded in the full range of human rights and the proactive engagement of States, companies, international organizations, civil society and the media';[9] against this context, State authorities are recommended to 'prioritize non-legal measures of countering disinformation and propaganda.'[10]

Standards concerning jurisdiction

Questions concerning 'Internet jurisdiction' have been widely discussed over the last few years and still prove widely contentious. For lack of more generally agreed principles,'country-of-receipt' rules, where a country purports to regulate foreign media outlets on the basis of 'mere accessibility' rules, are widely considered to run against the best practice of jurisdictional self-restraint and can easily either prove ineffective (when the country of receipt lacks the means to effectively exercise jurisdiction) or push international media outlets to comply with the most restrictive set of national rules (the so-called 'race to the bottom' effect), to the detriment of the freedom to impart and receive information. The Council of Europe recommends that, in principle, States should 'only exercise jurisdiction over foreign materials that are not illegal under international law in limited circumstances, notably when there is a clear and close nexus between the materials or the disseminator and the state taking action'[11] and thus make limited use of targeting tests, identifying, through precise connecting factors, specific cases when a media outlet has specifically targeted its activities at the domestic audience.

For example, with regard to the audio-visual media sector (both broadcasting and on-line), a particularly suitable connecting factor is the place where editorial decisions are taken (meaning that a service provider, even if formally established and registered abroad, exercises its editorial discretion within the territory of a country, then it ought to comply with local law); alternatively, other possible connecting factors are the place where a media outlet's head office or significant workforce are based, as provided under EU law.[12] Laws should not aim to subject programmes broadcast or distributed online by foreign media companies to domestic standards

[6] Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on 'Disinformation and freedom of opinion and expression', A/HRC/47/25, 2021, para. 55.
[7] Ibid., para. 89.
[8] Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on 'Disinformation and freedom of opinion and expression during armed conflict', A/77/288, 2022, para. 113.
[9] Report on 'Disinformation and freedom of opinion and expression', para. 87.
[10] Report on 'Disinformation and freedom of opinion and expression during armed conflict', para. 115.
[11] Council of Europe Commissioner for Human Rights, 'The rule of law on the Internet and in the wider digital world', Executive summary and Commissioner's recommendations, 2014, p. 23.
[12] See Directive (EU) 2018/1808 of 14 November 2018 (Audiovisual Media Services Directive), Art 2.

on the sole basis of their receivability: country-of-receipt rules normally provide strong targeting tests, such as for instance 'advertisement or other promotions specifically aiming at customers in its territory, the main language of the service or the existence of content or commercial communications aiming specifically at the audience in the Member State of reception'.[13]

Standards concerning the notions of media, platforms and journalism

Over the last couple of decades, the nature and shape of the media and communication industries have changed immensely as a result of technological development and usage habits at the global level. Digitalisation and convergence have progressively blurred the lines between industries that used to seem distinct and easily categorised as different mass media, such as the printed press, broadcasting and communication networks.

Policy makers have tried to constantly adjust the relevant principles and guidance so as to best capture the evolving nature of technologies while upholding the fundamental freedoms at stake. In fact, as a general principle to guide regulators and law-makers, the General Assembly of the United Nations has advised that 'the same rights that people have offline must also be protected online, in particular freedom of expression, which is applicable regardless of frontiers and through any media of one's choice.'[14]

Digitalisation and convergence have challenged the basic assumption that different mass communication technologies would need radically different policy and regulatory frameworks. However, the Council of Europe in its Recommendation on a new notion of media has advised law-making authorities to take a 'graduated and differentiated' approach aimed at encompassing all the different actors and entities involved in the production and dissemination on information. In order to determine the most suitable policy and regulatory frameworks capable of offering a 'flexible response, tailored to a concrete case (namely differentiated) and graduated for the purpose', authorities are advised to take into consideration a range of different criteria and indicators, such as an entity's intent to act as media, its purpose and underlying objectives, whether and to what extent it exercises editorial control over the content it distributes, whether and to what extent it abides by professional standards, its outreach and dissemination, whether and to what extent the public at large perceives it as a media outlet.[15]

As technological and industry convergence has continued to expand over the years, the role of digital platforms has radically changed. In a recent Recommendation on principles for media and communication governance, the Committee of Ministers of the Council of Europe has acknowledged the scale and impact of technological development on how media content is distributed: 'media have become heavily dependent on platforms, with their content no longer being distributed exclusively through printed products, broadcasts, websites and media apps but also through the websites and apps of platforms'.[16] In turn, this new state of the art has brought about a new understanding of the powers and responsibilities of on-line platforms: '[p]latforms are not neutral but have assumed an active curatorial or editorial role, including through the use of algorithmic systems, in the dissemination of content produced by the media

[13] Audiovisual Media Services Directive, Rec. 38.
[14] Resolution adopted by the Human Rights Council, 'The promotion, protection and enjoyment of human rights on the Internet', A/HRC/20/L.13, 2012, para. 1.
[15] See Recommendation CM/Rec(2011)7 of the Committee of Ministers to member states on a new notion of media, adopted by the Committee of Ministers on 21 September 2011.

[16] See Recommendation CM/Rec(2022)11 of the Committee of Ministers to member States on principles for media and communication governance, adopted by the Committee of Ministers on 6 April 2022, Preamble.

and by others, and thus have a huge impact on the way people perceive the world and are exposed to other information and ideas.'[17] Thus, the policy and governance of the media and communication sector has to adapt to this new reality, following the basic principle of the equivalence between off-line and on-line fundamental freedoms: 'despite this structural transformation of the public sphere brought about by digitalisation, the aims of media and communication governance have not changed, but … to be able to continue realising them, media and communication governance needs to be modernised to cover both the media and platforms, as they both play an essential role in facilitating communication in the public sphere, and further realising that States cannot and should not address all challenges alone but that those in the private sector should bear responsibility as well.'[18]

The same Recommendation has introduced a new and more specific definition for digital platforms, i.e. 'providers of digital services that connect participants in multisided markets, set the rules for such interactions and make use of algorithmic systems to collect and analyse data and personalise their services': this definition expressly includes both social networks and search engines, alongside news aggregators and video-sharing services. Platforms' self-regulatory efforts, normally deployed through internal terms of service, have been explicitly acknowledged as 'private ordering initiatives', defined as 'initiatives by individual private sector organisations to develop and enforce rules that may not only apply within their organisation (an organisation's internal editorial guidelines, for example) but potentially also affect users of their services (such as a platform's terms-of-service agreement or what are known as community standards)'.[19]Considered together, the recommendations and guidance on digital media stem from the fundamental observation that new entities in the current media ecosystem, despite structural differences from traditional media outlets, play a role that is complementary or even comparable to that of traditional mass media like broadcasters or the printed press. Therefore, different actors should bear obligations apportioned according 'to their specific functions in the media process and their potential impact'.[20] At a more general level, it is advisable to refrain from adapting older regulatory approaches to newer technologies. As recommended by the representatives of the international mechanisms for promoting freedom of expression, '[g]reater attention should be given to developing alternative, tailored approaches, which are adapted to the unique characteristics of the Internet, for responding to illegal content.'[21]

In relation to the notion of journalism, international and regional approaches have similarly developed towards a more flexible notion, so as to best capture the increased ability to disseminate content of public interest that private individuals and civil society organisations have today as a result of digital technologies. The UN Special Rapporteur considers that 'neither the concept of a journalist nor the practice of journalism is limited to those employed by news publishers', and therefore legal protection should be granted having in mind 'the nature of the content and its public interest function' more than 'the professional designation of the individual'.[22]

---

[17] Ibid.
[18] Ibid.
[19] Ibid.,, Appendix – Scope and Definitions, para. 4.
[20] Recommendation on a new notion of media, para. 2.
[21] Declaration signed by the UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and ACHPR Special Rapporteur on Freedom of Expression and Access to Information on 1 June 2011, 'Joint Declaration on Freedom of Expression and the Internet', para. 1.d.
[22] Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on 'Reinforcing media freedom and the safety of journalists in the digital age', A/HRC/50/29, 2022, paras. 15-16.

On a similar line, the European Court of Human Rights has stressed on multiple occasions that 'the function of creating various platforms for public debate is not limited to the press but may also be exercised by, among others, non-governmental organisations, whose activities are an essential element of informed public debate'.[23] Provided that they exercise a role comparable to traditional media in the public sphere, such as for instance that of a public watchdog, a variety of other entities can deserve similar protection under the law such as NGOs,[24] campaigners[25] or academics[26] on the basis that they play a role of similar importance to that of the press. Most relevantly to the issue under consideration, the Court stated that 'given the important role played by the Internet in enhancing the public's access to news and facilitating the dissemination of information, the function of bloggers and popular users of social media may be also assimilated to that of "public watchdogs" insofar as the protection afforded by Article 10 is concerned.'[27]

Standards concerning independence of media operators, intermediaries and regulatory authorities

The guidance from the different regional and international authorities underlines the need for striking a suitable balance between extended reliance on top-down regulation, traditionally associated with the risk of stifling media freedom, and appropriate self- or co-regulatory mechanisms while at the same time, also being wary of excessively relying on internal standards expressed in private companies' terms of service. Specifically on this latter point, the UN Special Rapporteur has recommended that States 'should avoid delegating responsibility to companies as adjudicators of content, which empowers corporate judgment over human rights values to the detriment of users'.[28]

It is thus advisable that law-making states authorities and the private sector collaborate towards developing co-regulatory frameworks. The Council of Europe Recommendation on the roles and responsibilities of internet intermediaries requires States to 'encourage appropriate self-regulatory frameworks or the development of co-regulatory mechanisms'[29]. Much along the same lines, in 2011 the UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression and the ACHPR Special Rapporteur on Freedom of Expression and Access to Information in their Joint Declaration on Freedom of Expression and the Internet advised that '[s]elf-regulation can be an effective tool in redressing harmful speech, and should be promoted'.[30] At the OSCE level, several declarations have variably recognised that 'independent media are essential to a free and open society',[31] and encouraged the adoption of

[23] ECtHR, *Magyar Helsinki Bizottság v. Hungary*, App. no. 18030/11, 8 November 2016, para. 166.
[24] ECtHR, *Társaság a Szabadságjogokért v. Hungary*, App. no. 37374/05, 14 April 2009; *Youth Initiative for Human Rights v. Serbia*, App. no. 48135/06, 25 June 2013.
[25] ECtHR, *Steel and Morris v. the United Kingdom*, App. no. 68416/01, 15 February 2005.
[26] ECtHR, *Başkaya and Okçuoğlu v. Turkey*, App. nos. 23536/94 and 24408/94, 8 July 1999.
[27] ECtHR, *Magyar Helsinki Bizottság*, para. 168.
[28] Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/38/35, 2018, para. 68.
[29] See Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, adopted by the Committee of Ministers on 7 March 2018, para. 1.3.10.
[30] Joint Declaration on Freedom of Expression and the Internet, para. 1.e.
[31] Document of the Moscow Meeting of the Conference on the Human Dimension of the CSCE, 1991; CSCE Budapest Document 1994 'Towards a Genuine Partnership in a New Era'; Decision No. 193. Establishment of the Office of the OSCE Representative on Freedom of the Media, Mandate of the OSCE Representative on

self-regulation and other mechanisms possibly capable of improving the independence, as well as overall quality, of media and journalistic outlets.[32]

While the principles above have been drafted with specific reference to media operators in the more classic sense, their applicability to online intermediaries should be undisputed to the extent that restrictions to their own independence might well reverberate on restrictions to the free circulation of media content. On this account, the Council of Europe has recognised that in today's fluid media environment the lines between different types of operators are often blurred[33] and therefore 'intermediaries and auxiliaries should be free from pressure or influence intended to bear on media, its independence or its editorial decisions'.[34] Following on the acknowledgment of this fundamental principle, the Committee of Ministers has also gone on to recommend that States 'encourage appropriate self-regulatory frameworks or the development of co-regulatory mechanisms' for online intermediaries as well, 'taking due account of [their] role … in providing services of public value and facilitating public discourse and democratic debate, as protected by Article 10 of the Convention.'[35] Similar considerations apply to search engines as well, in fact the Council of Europe has similarly invited State authorities to 'promote transparent self- and co-regulatory mechanisms for search engines, in particular with regard to the accessibility of content declared illegal by a court or competent authority, as well as of harmful content'.[36]

Guidance from the Council of Europe recommends that state authorities engage constructively with digital platforms through a 'clear, target-driven and accountable approach';[37] to any extent, the governance of the media and communication industry should align with the international standards for human rights: '[m]edia and communication governance should aim to promote human rights and fundamental freedoms in communication as they are essential for the functioning of democratic societies. This includes guaranteeing the widest possible exercise of these freedoms and limiting restrictions to what is necessary for the efficient protection of Council of Europe standards and values while encouraging industry self-regulation and private ordering initiatives. It also entails aligning rules for the offline and online environments, while guaranteeing free and independent media, platforms and communication.'[38]

Concerning instead the governance level, the independence of regulatory agencies for the media industry has been a long-standing standard established by international and regional institutions, with specific regard to the broadcasting sector. The Council of Ministers' Recommendation on the independence and functions of regulatory authorities for the broadcasting sector, stressing how 'regulatory authorities should have the power to adopt

Freedom of the Media (PC.DEC/193); Declaration on the Occasion of the 60th Anniversary of the Universal Declaration of Human Rights (MC.DOC/2/08).

[32] See OSCE, Decision No. 13/06, Combating Intolerance and Discrimination and Promoting Mutual Respect and Understanding (MC.DEC/13/06), 2006, para. 9.

[33] Recommendation on a new notion of media, para. 47.

[34] Ibid., para. 75.

[35] Recommendation on the roles and responsibilities of internet intermediaries, para. 1.3.10.

[36] Recommendation CM/Rec(2012)3 of the Committee of Ministers to member States on the protection of human rights with regard to search engines, adopted by the Committee of Ministers on 4 April 2012,, para. 8.

[37] Steering Committee for Media and Information Society (CDMSI), 'Content Moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', Guidance Note, 2021, p. 41.

[38] Recommendation on principles for media and communication governance, Appendix – Procedural principles for media and communication governance, para. 6.

regulations and guidelines concerning broadcasting activities'[39] within the context of rules and procedures that 'clearly affirm and protect their independence.'[40]

Although not clearly stated at the time of this Recommendation, it follows logically from the premise above (that digital platforms now perform functions complementary or comparable to that of traditional media, including broadcasters) that this principle should apply to converged regulatory authorities when their remit includes digital platforms.

Standards concerning monitoring and content obligations

Regional and international standards concerning private entities' responsibilities to monitor can be generally distinguished in two aspects. On the one hand, a first reason of concern is the kind of duties that might be legitimately imposed on private entities, in particular whether they should be tasked by public authorities with specific responsibilities for monitoring the presence on their service of content that could be variably described as illegal or harmful. On the other hand, a second reason of concern is what substantive norms and standards private entities should apply when fulfilling such duties, particular with respect to actions that impact directly on information distributed through their service, such as for instance removals or content moderation.

In regard to the first aspect, the Council of Europe Recommendation on the roles and responsibilities of Internet intermediaries advises that 'State authorities should avoid any action that may lead to general content monitoring'.[41] The UN Rapporteur took a similar stance when advised that 'States and intergovernmental organizations should refrain from establishing laws or arrangements that would require the "proactive" monitoring or filtering of content'.[42] Further to this, the same Council of Europe Recommendation also advises that 'State authorities should obtain an order by a judicial authority or other independent administrative authority, whose decisions are subject to judicial review, when demanding intermediaries to restrict access to content. This does not apply in cases concerning content that is illegal irrespective of context, such as content involving child sexual abuse material, or in cases where expedited measures are required in accordance with the conditions prescribed in Article 10 of the Convention.'[43]

In regard to the second aspect, a particular reason of concern has been the inchoate tension between digital intermediaries' global reach and States' jurisdiction naturally restricted to their own national boundaries, which has variably resulted in either an increasing reliance on private entities' own terms of service or self-regulatory codes of conduct/practice to set applicable standards, or on attempts from State authorities to mobilise private intermediaries to enforce national rules often beyond the boundaries of their own domestic jurisdiction.

Guidance and standards from international and regional authorities navigate this tension by drawing a delicate balance. On the one hand, the UN Special Rapporteur has expressed concerns for the often vague wording of state laws that aim to regulate content standards and impose content removal (or similar) obligations: platforms have been considered 'ill equipped

[39] Recommendation Rec(2000)23 of the Committee of Ministers to member states on the independence and functions of regulatory authorities for the broadcasting sector, adopted by the Committee of Ministers on 20 December 2000, para. 12.
[40] Ibid., para. 1.
[41] Recommendation on the roles and responsibilities of internet intermediaries, para. 1.3.5.
[42] Report A/HRC/38/35, para. 67.
[43] Recommendation on the roles and responsibilities of internet intermediaries, para. 1.3.2.

to make determinations of content illegality' and the lack of clear and precise enough guidance in a law often results in over-removals and self-censorship from users.[44]

On the other hand, the privatisation of content standards through platforms' terms of service has also been met with concern; vague wording and inconsistent application across the different platforms make it difficult for users to regulate their conduct accordingly and foresee the consequences thereof.[45] This increasing reliance on terms of service means in practice that states can leverage private power to remove 'a wide range of legitimate but (perhaps to some audiences) "uncomfortable" expressions.'[46]

Digital platforms are thus at the same time exposed to States' pressure to enforce domestic standards and in a unique position to selectively remove content, including lawful, but to varying degrees unwelcome, information. To resolve this tension, the Council of Europe has indicated that, when private companies are called to collaborate in the development of content standards, States bear 'a variety of positive and negative obligations in this context. They must create sufficiently developed regulatory frameworks for content moderation that upholds the exercise and enjoyment of human rights of internet users, including victims of illegal content. States must protect the rights to freedom of expression, privacy, freedom of assembly and association, equality and non-discrimination, the right to an effective remedy and other human rights of everyone within their jurisdiction when these rights are affected by content moderation.'[47] More specifically, States ought to ensure that any measures that platforms have to undertake are predictable, do not result in overcompliance or are enforced discriminatorily, and clearly indicate the degree of state engagement in each action taken by private actors;[48] take into account the different sizes and scales of platforms, avoid imposing disproportionate obligations or delegating decision-making responsibilities 'to the detriment of democratically legitimated approaches'.[49] Predictability ('States should ensure that in all cases legal obligations and responsibilities, as well as operational roles and accountability requirements are clearly defined'[50]) and proportionality ('If internet intermediaries are to be held liable for failing to remove illegal content, the rules concerning "knowledge" triggering that liability must be clear and proportionate, as must the rules prohibiting the content in question;'[51] furthermore, 'States must ensure an appropriate balance of incentives for internet intermediaries and avoid regulation incentivising them to impose disproportionate restrictions. This can happen, for example, as a result of intermediary liability rules that are either too stringent or too vague. This is particularly relevant for content that is legal but possibly undesirable in a democratic society, where it is recognised that human rights must also be upheld'[52]) are key principles that platforms' standards must uphold at all times.

Against this background, the governance system of the media and communication industries should ensure that all actors involved 'comply with content obligations in accordance with Article 10 of the Convention and with professional standards. This includes clearly defining illegal content and addressing legal but harmful content, the possibility of other public interest content requirements, effective measures against violations of content standards, and

[44] Report of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/32/38, 2016, paras 39 and 44.
[45] Ibid., para. 52.
[46] Ibid., paras. 52-53.
[47] Council of Europe, 'Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', Guidance note adopted by the Steering Committee for Media and Information Society (CDMSI), 2021, para. 12.
[48] Ibid., para. 14.
[49] Ibid., para. 21.
[50] Ibid., para. 24.
[51] Ibid., para. 23.
[52] Ibid., para. 28.

redress mechanisms.'[53] To any extent, private entities should 'recognize that the authoritative global standard for ensuring freedom of expression on their platforms is human rights law, not the varying laws of States or their own private interests, and they should re-evaluate their content standards accordingly.'[54]Following the recent Council of Europe's guidance, ensuring the transparency of the system ought to comprise two different but interrelated aspect: clarity of the rules imposed in the terms of service needs to be paired with clear and detailed information about the different measures taken, such as what content has been removed and on what ground, whether it was removed on the basis of company terms of service or for constituting a statutory offence. Best practices recommended by the Council of Europe include making available specific break-downs of complaints received, specific reasons for any decisions made, especially when content is removed. When multiple companies or platforms provide data, as in this case, the methodology utilised should be standardised and machine-readable. The availability of such kind of data is particularly key to ensure that the effectiveness of any regulatory approach can be assessed on an ongoing basis; objectives need to be established clearly and unequivocally, so as to avoid industry players 'gaming' the system or pursuing misguided goals, as it happens for instance when speed of removals is prioritised over accuracy. Data that should be collected in a granular manner include: 'the specific reason for the removal of the content (terms of service or law), if data is stored in relation to potentially criminal material, if this [data] is available to law enforcement authorities on demand, if or how often this data was requested by law enforcement authorities, how often content was not immediately removed in order to avoid interfering with law enforcement investigations, other reasons for delays in takedowns, speed of takedowns/blocks.'[55] In respect to end-to-end encryption technology, the Council of Europe Commissioner on Human Rights expressed concerns over legal requirements to use detection tools capable of circumventing end-to-end encryption: 'Effective end-to-end encryption is indispensable if we want to protect the security of communications for everyone on a network. We need encryption to ensure that no-one, including the platform provider, can read or alter messages, and to preserve the confidentiality between the sender and the recipient. Encryption is therefore indispensable for the effective protection of the right to privacy, freedom of expression, and many other human rights.'[56] A few years earlier, the UN Commissioner on Human Rights had expressed a very similar concern: 'Encryption and anonymity are needed as enablers of both freedom of expression and opinion, and the right to privacy. It is neither fanciful nor an exaggeration to say that, without encryption tools, lives may be endangered.'[57]

A growing trend at the international level is the use of automated detection tools by platforms to detect and intervene on illegal or harmful content. Such technologies can offer a valuable resource to help platforms act effectively and promptly on their obligations; platforms thus face increasing pressure to adopt them, and at times are even imposed specific legal

[53] Recommendation on principles for media and communication governance, Appendix – Procedural principles for media and communication governance, para. 10.
[54] Report A/HRC/38/35, para. 70.
[55] 'Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', p. 44. For an example of reporting made useless by the lack of granularity, see in the same page: 'YouTube's transparency report under the German Network Enforcement Law includes a category called "terrorist or unconstitutional content" which mixes terrorist content with breaches of provisions of the German Criminal Code generally, but not totally, unrelated to terrorism, such as use of symbols of unconstitutional organisations (article 86a of the Criminal Code) and certain types of forgery (article 269 of the Criminal Code).'
[56] Opening intervention at European Digital Rights Association (EDRi) event: "Encryption in the age of surveillance", by Dunja Mijatović, Council of Europe Commissioner for Human Rights, 2023.
[57] Office of the High Commissioner for Human Rights, Press release: 'Apple-FBI case could have serious global ramifications for human rights: Zeid', 2016.

obligations in this sense. However, this practice also raises crucial concerns for a variety of reasons, more specifically for these technologies' potential of imposing prior restraints to freedom of expression, their inherent lack of transparency and ineffectiveness in understanding context,[58] which in turn leads them to being 'prone to overbroad application of the rules they seek to impose,'[59] and their high risk of bias.[60] For these reasons, guidance from the OSCE recognises that States are the ultimate 'duty-bearers under international human rights law and hold a positive obligation to protect human rights from interference by others, including private actors or individuals' and therefore recommends that they develop 'a human rights policy with emphasis on salient human rights issues, including freedom of expression, freedom of the media, privacy, non-discrimination, and right to life, liberty and security.'[61] Contextually, States are urged not to introduce legal requirements for platforms to deploy proactive automated tools to detect and identify illegal or harmful content; to provide clear and detailed guidance, normally via their judicial authorities, on what is considered illegal content under the applicable legislative framework and to differentiate between various categories of illegal content; to legally mandate human rights due diligence for algorithmic content moderation and data-harvesting business models.[62] Further recommendations concern requirements for algorithmic transparency: States should oblige internet intermediaries to provide documentation on the AI tools they deploy for content moderation, in such a way that is understandable and accessible for all users; to provide documentation concerning content-specific models adopted; to adopt diverse datasets, based on diverse attributes, so as to contrast the risk of algorithmic bias and discrimination.[63] Further obligations that States should impose on service providers concern notifying users when automated systems are used to moderate third-party content, allowing users to opt-out of automated decision-making, providing meaningful transparency reporting on actions taken in response to potentially lawful but harmful content, and publishing the number of reports of abusive or harmful conduct received each year.[64]

Finally, the UN Rapporteur has recommended that States 'refrain from adopting models of regulation where government agencies, rather than judicial authorities, become the arbiters of lawful expression'.[65] On the same point, the Council of Europe accepts that decisions to restrict access to content might as well come from 'independent administrative authorit[ies] provided that such decisions are subject to judicial review, with the exception of content that is illegal 'irrespective of context' (such as content involving child sexual abuse material) or 'in cases where expedited measures are required in accordance with the conditions prescribed in Article 10 of the Convention'.[66] Overall, the Council of Europe states that 'States have the ultimate obligation to protect human rights and fundamental freedoms in the digital environment' and advises that '[a]ll regulatory frameworks, including self- or co-regulatory approaches, should include effective oversight mechanisms to comply with that obligation and be accompanied by appropriate redress opportunities'.[67]

[58] See OSCE, 'Spotlight on Artificial Intelligence and Freedom of Expression A Policy Manual', 2022, p. 25.
[59] Ibid., p. 17.
[60] Ibid., p. 35.
[61] Ibid., pp. 45-46.
[62] Ibid.
[63] Ibid., 38-39.
[64] Ibid., 39-45.
[65] Report A/HRC/38/35, para. 68.
[66] Recommendation on the roles and responsibilities of internet intermediaries, para 1.3.2.
[67] Recommendation on the roles and responsibilities of internet intermediaries, para. 1.1.3.

IV. Overview and analysis of the proposed legal reform

Scope and jurisdiction

The Act has a very broad scope in regard to both the kinds of service covered by its provisions and the geographical application.

In regard to the first perspective, the Act aims to regulate platforms mainly dedicated to providing 'user-to-user services' (defined in Section 2 as any 'internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, may be encountered by another user, or other users, of the service') as well as search engines.

The Act foresees a graduated approach, with service providers divided into three categories depending on their number of users, what functionalities are offered, and the level of risk of harm posed by the content that the service helps to disseminate. This approach has also been deployed by other recent legal frameworks at the comparative level, most recently the EU's Digital Services Act which introduced the two categories of 'very large online platform' and 'very large online search engine'. The DSA, however, also provides a precise definition of the boundaries of such categories (a platform for instance would qualify as 'very large' if its number of average monthly users is 10% or more of total consumers across the EU) whereas the Act provides that the Secretary of State will issue regulations to determine the threshold for each category, and will subsequently maintain the power to amend the original regulations in the future, following research carried out by OFCOM. Considering in particular how service providers will bear more or less intensive duties depending on which category they belong to, the provision fails to offer enough clarity and foreseeability as to the exact nature and intensity of the provisions that each service provider will have to comply with.

In discharging this function, the Secretary of State and OFCOM will have to strike a balance between keeping a sufficient degree of flexibility for the law to adapt to the changing structure of the market and/or possible technological development, and the legitimate expectation, from the regulated entities, of an adequate degree of predictability and consistency. The Secretary of State and OFCOM, while periodically revising the thresholds of each category, could consider providing a timeline for periodic revisions, so as to guarantee that regulatory requirements will remain unchanged at least for a set period and changes would happen at a foreseeable time. OFCOM could offer support and advice, when needed, to services in the transition from one category to another.

In regard to the second perspective, the Act aims to have extra-territorial application and, pursuant to Section 4[68], it is deemed for application to service providers with 'links' to the UK such as a significant number of users in the UK, being targeted towards UK users, or being capable of being used by individuals in the UK and is likely to cause significant harm to individuals in this country. Further to this, Section 165 expressly provides for the extra-territorial application of the Act, with regard to the offences of sending false communications, sending threatening communications, and sending or showing flashing images, if committed from outside the UK by a person who habitually resides in England or Wales.

Like in the case of the different categories illustrated above, the provision in Sec. 4 fails to identify with enough precision the boundaries of its own scope: in this case, it fails to quantify the concept of a 'significant' number of UK-based users and is therefore likely to give rise to situations of dubious application to foreign service providers, to the detriment of legal certainty.

[68] Sec. 3 of HL Bill 87(Rev).

Therefore, given the negative consequences of such uncertainty on the free flow of information including on the ability to seek, receive and impart information and ideas regardless of frontiers, OFCOM should provide clear indications of the circumstances under which a foreign service provider would be deemed to have sufficient links to the UK to fall under the scope of this provision.

The 'news exemption' and respect for media freedom

Throughout the Act, several provisions are covered by what appears as a 'news exemption' which aims to exclude the applicability of such norms to content originating from news media outlets. Relevant provisions include: a duty to notify news publishers before taking down, restricting users' access or taking other action (such as for instance curation) in relation to content, and allow a reasonable time for the news publisher to make representations, and immunity from the offence of false communications provided in Section 160.

The notion of journalistic content is described in Section 15[69] in relation to a duty for Category 1 service providers to take into account the 'importance of the free expression of journalistic content' while making decisions about taking down content or restricting access to it, and to provide a dedicated and expedited complaints procedure in relation to such decisions.

The material scope of the 'news exemption' includes any content defined as 'generated directly on the service by a user of the service that is a recognised news publisher' and complete reproductions or recordings thereof, or a link to any such content (Section 49(9)-(10), as well as material consisting of news, information or opinion about current affairs, or gossip about celebrities, other public figures or other persons in the news, published by any means (including by broadcasting, and subject to a code of standards published by either an independent regulator or by the entity itself) (Section 50(5)).

The personal scope of the Act defines 'recognised news publishers' as the BBC, a licensed broadcaster, or an entity that meets all of a series of conditions (having as its principal purpose the publication of news-related material, being made of more than one person, exercising editorial control, publishing news in the course of business, including on a non-for-profit basis, being subject to a standards code, having internal procedures in place for handling and resolving complaints) and, from a jurisdictional perspective, has a registered office or a business address in the UK, the person with legal responsibility for material published by it is in the UK.

The definition is evidently focused on institutional media outlets and runs against both the general principles of a graduated and differentiated approach recommended by the Council of Europe and the functional approach to defining journalism emerging from the case-law of the European Court of Human Rights. By contrast, the Act focuses on a range of formalistic indicators that, by capturing exclusively traditional and mainstream media outlets, are likely to leave out a series of other media actors that play an important role in providing information of public interest to the public. In fact, over the last few years independent media outlets[70]

[69] Sec. 14 in HL Bill 87(Rev).

[70] The Reuters Institute noted, in its annual Report in 2022, that 'social networks have steadily replaced news websites as a primary source for younger audiences overall, with 39% of social natives (18–24s) across 12 markets now using social media as their main source of news … .Social natives are far more likely to access news using 'side-door' sources such as social media, aggregator sites, and search engines than older groups.' Reuters Institute Digital News Report 2022, p. 42.

(including individual content creators[71]) as well as civil society organisations[72] have increasingly become an integral part of news consumption routine for larger shares of the audience, thanks to the ability afforded by digital media to communicate directly with the public bypassing traditional gatekeepers. Most often, alternative outlets give voice and representation to more marginalised communities that struggle to find representation on larger and more institutionalised media.[73]Focusing narrowly on traditional, institutional media outlets would have adverse consequences on such a segment of the digital media industry that is increasingly becoming the main source of news for younger consumers and, consequently, on those consumers who rely on them to exercise their right to receive information.

With more specific regard to the material scope of the exemption, it seems worth noting that Section 13 provides for a generic duty to protect content of democratic importance, defined as content 'specifically intended to contribute to democratic political debate in the United Kingdom or a part or area of the United Kingdom'. Section 14 provides for a duty to protect news publisher content, while Section 15 provides for a duty to consider 'the importance of the free expression of journalistic content' before taking action such as taking it down or restrict users' access to it. The provision defines 'journalistic content' as either originating from a news publisher or as regulated user-generated content – thus seemingly and encouragingly acknowledging the possibility that any such material originates from non-professional entities; the content must also be 'generated for the purposes of journalism' and UK-linked. The provision seems however quite generic and does not define the methods or parameters by which service providers are expected to determine, in their own terms of service, whether content is of journalistic character, how its importance is taken into account, and how any decisions concerning the appropriate actions to be taken should be made. Further to this, the three provisions considered above apply exclusively to Category 1 service providers; but in lack of indication on what the contours of the category would be, it is not immediately evident why the same obligations should not apply to the other categories, in a proportionate manner.

In its final wording, the Act does not offer adequate protection to a sizeable part of the media industry – such as for instance freelance reporters or smaller independent outlets. Pursuant to the latest amendment to the Introduction passed in September 2023, any duties imposed with the Act ought to secure and protect 'users' rights to freedom of expression and privacy'. Therefore, it will be important that service providers and any authorities implementing the Act in the future interpret it in line with the principles established by the Council of Europe and the European Court of Human Rights, including Act by interpreting the notions of news publishers and journalistic content from a perspective that takes into account the role and relevance that content put forward by non-professional journalistic endeavours can have for the public debate and society at large, rather than by looking solely at the characteristics of the institutions publishing the content at stake, to the extent that is possible for the purposes of this legislation.

In this respect, it might be useful to observe that the Information Commissioner's Office, faced with a similar question as to how to interpret the notion of journalism in the context of the Data Protection Act, advised that the notion of journalism 'should be interpreted broadly'[74] from at least two different perspectives: in regard to the material scope, the notion

[71] See M. Riedl et Al., 'The Rise of Political Influencers—Perspectives on a Trend Towards Meaningful Content', Frontiers, vol. 6, 2021, available at: Frontiers | The Rise of Political Influencers—Perspectives on a Trend Towards Meaningful Content (frontiersin.org)

[72] See M. Powers, *NGOs as Newsmakers. The Changing Landscape of International News*, Columbia University Press, 2018.

[73] See, for example, G. Khan, 'Forced out from print and airwaves, news media in Venezuela shift to digital to survive', 14.3.2023, https://reutersinstitute.politics.ox.ac.uk/news/forced-out-print-and-airwaves-news-media-venezuela-shift-digital-survive.

[74] Information Commissioner's Office, 'Data protection and journalism: a guide for the media', 2014, p. 29.

can 'potentially cover almost all information collected or created as part of the day to day output of the press and broadcast media, and comparable online news or current affairs outlets';[75] in regard to its personal scope, 'individuals may be able to invoke the journalism exemption if they are posting information or ideas for public consumption online, even if they are not professional journalists and are not paid to do so'.[76] Building on case-law from European and domestic courts, the ICO recommends that the notion of journalism is interpreted dynamically taking into account elements such as: the purpose of the publication; how closely the activity resembles more traditional media activities; the content of the information and whether there is any public interest in it; the means by which the information was published; the extent to which it has been promoted to the public; and any restrictions on its use.[77]

Lack of independence of the regulator

The Act generally displays a strong emphasis on top-down regulation. Section 36 requires OFCOM to issue codes of practice 'describing measures recommended for the purpose of compliance' with duties set out in respect of illegal content or terrorism offences. The responsibility for drafting (and amending as necessary) such codes will lie with OFCOM, a regulator for communication services, in the first place and then with the Secretary of State, to whom the regulator has to submit the draft, and eventually the Parliament to which the Secretary of State must transmit the draft. OFCOM, the Secretary of State and the Parliament, all have extensive powers to determine the content of codes of conduct

OFCOM in fact is required to submit draft codes, or subsequent draft amendments, to the Secretary of State (Section 38) who, in turn, may direct OFCOM to make modifications (Section 39) on the basis of loosely defined grounds such as 'for reasons of public policy', 'for reasons of national security or public safety' or for reasons relating to 'relations with the government of a country outside the United Kingdom' – in fact so much vague as to leave the Secretary an almost unfettered discretion on the exercise of its regulatory power. Where the Secretary of State exercises such power, the draft code also ought to be submitted to the Parliament for scrutiny. Such extensive is the Parliament's power of scrutiny that it can prevent OFCOM from approving a draft by simply voting against it.

The grounds for the Secretary of State to direct OFCOM to modify a draft of the code under Section 39 are too broad and vague. Even more, the Secretary of State might issue a direction without providing a reason where it is considered that 'doing so would be against the interests of national security, public safety or relations with the government of a country outside the United Kingdom'.

As a point of reference, OFCOM is also tasked with setting and revising codes for the content of radio and television programmes (Section 319, Communications Act 2003). The Communication Act, however, does not envisage the Secretary of State or Parliament having any comparable role in regard to broadcasting codes. There is no apparent reason, however, as to why the regulator should enjoy different levels of independence when acting in respect of different industries. The principle underlying the regulatory model in the first place (that is, bestowing regulatory powers on an independent authority so as to protect the free flow of information from political and partisan influences) should still apply to the digital industry.
In the current Act, OFCOM itself is not adequately safeguarded from State authorities interfering with its own regulatory powers and does not seem to be granted a significant margin

[75] Ibid., p. 30.
[76] Ibid.
[77] Information Commissioner's Office, 'Draft journalism code of practice', 2021, p. 23.

to exercise its own discretion; in fact, para. 6 of Section 39 provides that OFCOM must 'comply' with the direction received from the Secretary of State and does not provide an alternative route for OFCOM to engage in a discussion with the Secretary. As a result, the overall mechanism effectively amounts to an indirect system for the political majority in power to discretionarily legislate on the flow of communications, with tangible risks for service users' rights to impart and receive information. It would be recommendable to consider ways to enhance protections of the regulator's independence from the executive power.

Monitoring duties

The Act establishes several duties for service providers in respect of different kinds of content. All providers of regulated user-to-user services must carry out risk assessments in relation to illegal content (as provided in Section 8), comply with reporting duties (Section 16) and set up complaint procedures (Section 17) and protect freedom of expression and privacy (Section 18(2)-(3)). In addition to those, providers of Category 1 services must also comply with duties to empower adult users (Section 12) and protect specific categories of content such as content of democratic importance (Section 13), news publisher content (Section 14), journalistic content (Section 15) and freedom of expression and privacy (Section 18(4)-(6)-(7)).

In regard to illegal content, Section 8 requires all regulated user-to-user services to carry out 'a suitable and sufficient' illegal content risk assessment, based on a range of indicators among which feature the possible use of algorithms by the service provider, and the likely ease, speed and width with which illegal content could be disseminated through the means of the service. Section 9 requires service providers to take 'proportionate' measures aimed at preventing users from encountering priority illegal content (defined in Section 53 as terrorism content, Child Sexual Exploitation and Abuse content, or other content that amounts to an offence specified in Schedule 7) while using the service, mitigating the risk of the service being used for committing or facilitating a priority offence, mitigating the risk of harm to individuals, as identified in the risk assessment carried out by the service provider, minimise the length of time for which any priority illegal content is present, and finally to 'swiftly' take down illegal content when alerted to its presence or having become aware of it 'in any other way'. In complying with such duties, the Act indicates that service providers would be expected to make use of such measures as the 'design of functionalities, algorithms and other features'; sub-section (7) mentions a further duty to 'include provisions in the terms of service giving information about any proactive technology used by a service for the purpose of compliance' with the duties mentioned above.

Overall, while the breadth and scope of such monitoring duties seems necessary to achieve the ambitious goals of this regulatory effort, the provisions seem to lack of the necessary degree of precision to allow the regulated entities to understand the extent of their obligations and the concrete risk of incurring into penalties. Some of the provisions above are formulated in quite generic language (a duty to take down illegal content 'swiftly', an expectation that service providers make use of 'design …, algorithms and other features'. The timing and the methods that service providers will need to recur to will need to be specified in greater detail or else uncertainty over the risk of penalties might offer an incentive to over-removals.

It would add to the proportionality and predictability of the obligations imposed onto the service providers, as well as ultimately to the overall guarantees of the users' rights to impart and receive information, if OFCOM could provide for more specific and detailed guidance on the nature of the specific actions expected from the service providers; on the nature

of the technology that such providers would be expected to deploy, also taking into consideration the different capacity of different regulated entities within the industry; and on the exact timing of such obligations. In providing such guidance, it would be ideal if OFCOM could operate in collaboration with industry players to ascertain the feasibility of different options and with a particular view of avoiding excessive regulatory burdens which could in turn lead to the Act risk of over-removals of content.

Section 59 establishes a duty to report, requiring that regulated user-to-user services to report all detected and unreported child sexual exploitation and abuse (CSEA) content present on the service or, in the case of search engines, on websites or databases capable of being searched by the search engine, to the National Crime Agency. Non-UK providers are similarly required to do the same, inasmuch as the content at stake is UK-linked (i.e. when the content was published, generated, uploaded or shared in the UK; the person suspected of committing the related offence is a UK national or based in the UK; or the victim child is based in the UK). Section 60 provides that the Secretary of State will make regulations providing about different circumstances concerning the reports (such as the information included, the format, the manner, the time frames and the records to be kept). Section 63 clarifies that providers become aware of CSEA content, and thus responsible for reporting it to the NCA, by simply becoming 'aware of [it], whether by means of the provider's systems or processes or as a result of another person alerting the provider'.

The duty under Section 59 appears potentially compatible with international standards in as much as CSEA content might be easily identified as such by service providers without excessive discretion in their judgement. At the same time, the circumstance that the Secretary of State might discretionarily determine the content of required reports could open ways for problematic disclosure of users' personal information, including in principle those not directly responsible. It would be preferable for the law to determine what exact information the Secretary of State should be able to gain access to.

Section 9 requires a provider, 'where [it] is alerted by a person to the presence of any illegal content, or becomes aware of it in any other way', to 'swiftly' take it down. Such obligations to remove unlawful content (so-called 'take down' obligations) are now in principle accepted in different jurisdictions – including in the EU with the recent Digital Services Act.

Particularly relevant in this respect is the power granted to OFCOM to impose economic contraventions and penalties in different forms – either single or on a daily rate basis (Section 118 and 125-127) – for a failure to fulfil specific obligations in respect to a wide series of subject matters (including illegal content risk assessment, illegal content, children's risk assessments, content of democratic importance, news publisher content, content of democratic importance, fraudulent advertising, content reporting among others). Pursuant to Section 120-121, OFCOM can also request regulated services to 'take steps' to comply with a notified requirement or remedy a failure to comply with it. Among such actions that OFCOM can require, Section 124 includes the deployment of any 'proactive technology' in regard to illegal content, children's online safety, fraudulent advertising to analyse user-generated content communicated publicly and the related metadata. Section 124 states that OFCOM must identify the content, or the parts of the service that include such content, without further specifications.

The requirement of deploying 'proactive technology' appears particularly broad and vague, most immediately regarding the type of technology that service providers would be expected to use. Furthermore, it appears that such an obligation could easily turn into a proactive monitoring obligation. It is concerning that, by including in legal frameworks generic references to unspecified technologies, new and increasingly more active and general forms of monitoring become progressively normalised, silently eroding existing safeguards for freedom of expression and the free flow of information. In fact, the rapid advancement of technology has already proved to have a decisive impact on shifting assessments and attitudes regarding

the nature of monitoring obligations and their compatibility with human rights and international legal standards. As an example, commentators have already noted how recent obligations to adopt proactive measures included in the EU Copyright Directive and an early draft of the Regulation on addressing the dissemination of terrorist content online would amount to 'an actual (and declared) derogation of the non-monitoring principle'.[78] However, while harnessing the potential benefits of new technologies to advance the safety of the online environment remains a valuable opportunity, it is key that recourse to technology-based monitoring solutions is predicated on the basis of its specificity rather than open-endedness. To that end, in the EU Regulation on addressing the dissemination of terrorist content online, following an amendment tabled by the EU Parliament, a provision originally requiring platforms to adopt 'proactive measures to protect their services against the dissemination of terrorist content' was changed so as to require the adoption of 'specific measures' in the final text.[79]

Furthermore, recourse to automated and proactive technologies cannot be a substitute for a clear and detailed indication of the nature of the information that service providers are expected to remove or block. The recent decision of the Court of Justice of the European Union in *Glawischnig-Piesczek v. Facebook*, which accepted that digital intermediaries can be expected to utilise technological tools to fulfil their monitoring duties,[80] in fact was largely dependent on the circumstance that the injunction requiring intermediaries to resort to proactive technologies clearly indicated the 'specific elements' that platforms would need in order to identify the content which has to be removed.

On the contrary, this Act contains no indication that OFCOM (or a judicial authority) would provide guidance to service providers on what technologies to deploy, or on what 'specific elements' such technologies would need to be deployed to look for. While the promise of new technological tools to help make the Internet a safer place remains by all means most intriguing, it is also important that concerns are minimised by ensuring that OFCOM collaborates routinely with service providers to identify the technologies that service providers would be expected to deploy, and the specific kinds of usage that would not unduly impinge on fundamental rights. Collaboration with industry stakeholders would be all the more important when OFCOM, pursuant to its authority under Sec. 124, includes in a confirmation decision a requirement to use any proactive technology.

Content removal obligations

The Act identifies certain specific categories of content and corresponding actions expected from service providers, in particular:

- Category 1 and Category 2A providers need to prevent individuals from encountering fraudulent advertisements, minimise the length of time that such advertisements are present on their services, and alert users to the presence thereof;

[78] J. Barata, 'Positive Intent Protections: Incorporating a Good Samaritan principle in the EU Digital Services Act', Center for Technology and Democracy 2020, https://cdt.org/wp-content/uploads/2020/07/2020-07-29-Positive-Intent-Protections-Good-Samaritan-principle-EU-Digital-Services-Act-FINAL.pdf.

[79] Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online, Art 5. On the nature of the measures required and the amendment tabled by the EU Parliament, see A. Kuczerawy, 'To Monitor or Not to Monitor? The Uncertain Future of Article 15 of the E-Commerce Directive', 2019, https://balkin.blogspot.com/2019/05/to-monitor-or-not-to-monitor-uncertain.html.

[80] CJEU (Third Chamber), *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, Case C-18/18, 3 October 2019.

- Two general groups of content removal obligations: one that applies in respect of adult users, and one that applies for children.
  - In regard to the first group, service providers are required to remove illegal content which in turn is defined as 'words, images, speech or sounds' which either fall under one of the categories defined as 'priority offences' or another offence within subsection (5) of Section 53. Priority offences are defined as falling under one of the following categories:
    - Terrorism offences amount to a series of conducts prohibited under the Terrorism Act 2000, such as inviting support for a proscribed organisation, expressing an opinion or belief supportive of such organisations, publishing an image of their uniform, providing weapons training, etc. Some of the offences originally provided in the Terrorism Act 2000 and indicated in this Act as priority offences do not seem to easily lend themselves to being committed through the means of the services offered by the entities which the Act purports to regulate, and should be therefore interpreted with a certain degree of elasticity for this purpose (for instance, Sec 16 of the Terrorism Act, expressly included as a priority offence in Sec 53(1)(g)-(h) of the Act, prohibits the possession of money or other property intended to be used, or reasonably suspected to, for the purpose of terrorism, and as such can hardly be construed as an offence committed through the 'possession, viewing or accessing' or the 'publication or dissemination' of the relevant content, as provided in Sec 58(3)(a)-(b), however the provision might be possibly interpreted as applying to on-line transactions made through a regulated service).
    - Child sexual exploitations and abuse (CSEA) offences amount to a series of conducts prohibited under the Obscene Publications Act 1959, the Protection of Children Act 1978, the Protection of Children (Northern Ireland) Order 1978, the Criminal Justice Act 1988, the Sexual Offences Act 2003, the Sexual Offences (Northern Ireland) Order 2008, the Coroners and Justice Act 2009, or the Serious Crime Act 2015. These various provisions all refer to different instances of possession of indecent images of children, or inciting a minor to engage in sexual activities of different nature.
    - Other priority offences amount to a series of conducts prohibited under several statues related respectively to: assisting suicide; threats to kill; public order offences, harassment, stalking and fear or provocation of violence; supply or offer to supply drugs and psychoactive substances; purchase or sale of firearms and other weapons; assisting illegal immigration; causing or inciting sexual exploitation; concealing, facilitating the acquisition, use or possession of proceeds of crime; fraud; offences related to regulated financial services.
    - Other, non-priority offences are defined as affecting one or more individuals (possibly also unintendedly, as the wording of subsection (5)(b) of Section 53 – 'the victim or intended victim' – seems to suggest) and created by either the same Act under analysis, or any other Act, Order in Council, order, rules or regulations made by a Secretary of State or Minister, or devolved subordinate legislation.
  - In regard to the second group, i.e. user-to-user services likely to be accessed by children, Sections 10 and 11 provide for risk assessment and safety duties construed similarly to those in Sections 8 and 9, but without reference to the illegal nature of such content. Alongside with measures to prevent children accessing content that is harmful to them (e.g. age verification), Section 11(5)(e) mentions content moderation, 'including taking down content', among the measures that service providers would be

expected to deploy. In its latest version, the Act defines 'Primary priority content that is harmful to children' as a series of specified content including pornographic content, content encouraging suicide, or self-injury, eating disorders, in a variety of textual or graphic formats. 'Priority content that is harmful to children' is defined as a series of content such as abusive content targeting one or more protected characteristics, incites hatred against certain protected characteristics, encourages serious violence against a person, is of a bullying nature, depicts real or realistic serious violence or injury against a person, an animal or a fictional creature, encourages dangerous challenges or stunts, ingesting or inhaling harmful substances. A previous version of the Act delegated to the Secretary of State the authority to determine the exact nature of such categories of content; these new amendments are certainly welcome in that they offer more precision and predictability with regard to the service providers' obligations. However, the insertion of new clause after Sec. 191, which gives the power to amend such sections to the Secretary of State, might in turn undermine the benefits of the new and more precise provisions.

- Sections 160-166 provide for 'communications offences'. These include:
  - A false communications offence (Section 160), consisting of sending information known by the sender to be false, intended to cause non-trivial psychological or physical harm 'to a likely audience' (i.e. an individual whom would be reasonable to foresee encountering the message, either directly or as a result of it being subsequently forwarded or shared, irrespectively of whether the sender had originally intended that person to be harmed) without a reasonable excuse. Pursuant to Section 160(5), 'A person who commits an offence under this section is liable on summary conviction to imprisonment for a term not exceeding the maximum term for summary offences or a fine (or both)'. News publishers and licensed broadcasters are exempted from this provision (Section 161).
  - A threatening communications offence (Section 162), consisting of sending a message conveying 'a threat of death or serious harm' (i.e. 'grievous bodily harm' as defined in the Offences against the Person Act 1861, rape, assault by penetration as defined in the Sexual Offences Act 2003, or serious financial loss), which at the time was expressly or recklessly intended to cause the receiver a fear that the threat would be carried out.
  - Offences of sending or showing flashing images electronically (Section 164), consisting of sending flashing images, either to an indefinite receiver if the sender could have foreseen that an individual with epilepsy would have been reasonably likely to see them and intended to cause harm to such individual, or to a specific receiver whom the sender knew or suspected to be an individual with epilepsy and to whom the sender wanted to cause harm.
  - The offence of encouraging or assisting serious self-harm.
  - The offences of sending photograph or film of genitals; sharing or threatening to share intimate photograph or film (Sec. 167).

Overall, the Act imposes a broad and wide-ranging list of types of content on which service providers are expected to intervene. In many cases, there appears to be a basis in existing statutory provisions, while the Act also creates a few new offences (such as the false communications offence, the threatening communications offence, sending or showing flashing images, and criminalising assisting or encouraging self-harm online). It remains key, however, that any such provisions are interpreted as narrowly as possible, according to an equally compulsory principle of proportionality repeatedly and consistently affirmed by the European Court of Human Rights in its case-law. This especially concerns those provisions

whose wording could more easily lend itself to overbroad and expansive interpretation (e.g. 'public order offences' or 'assisting illegal immigration').

The category of 'false communication offences' appears more problematic from this perspective. While the rampant spread of harmful disinformation is certainly an urgent concern for public authorities, private entities, and other stakeholders to counter together, the notion of 'false communication offence', as worded in the Act, is a newly introduced offence with undefined contours. In fact, the provision fails to determine what specific harm a communication should be capable to cause in order to fall under this provision; and why, if harmful anyway, such communications would also need to be 'false' in order to be restricted. The possibility of imposing prison sentences for such offences is likely to be incompatible with international standards and recommendations. It should be considered introducing a more stringent definition of the category of false communication specifying in detail what specific harm such communications should attain or should be likely to attain in order to warrant regulatory attention, without resorting to criminal sanctions; service providers could be required to curate and moderate content so as to restrict the availability of content deemed as harmful.

Service providers' own terms of service and reporting obligations

Section 64 indicates that service providers can take down content, restrict users' access to information, or suspend or ban users 'in accordance with the terms of service'. Section 65 requires providers to use 'proportionate systems and processes' to this purpose, and the terms of service should be 'clear and accessible', 'written in sufficient detail to enable users to be reasonably certain' of whether action would be taken, and applied consistently. Section 66 provides that OFCOM will produce guidance for providers.

There is little clarity, however, on the circumstances under which platforms could deploy their own terms of service as the basis for removal decisions. As explained in the guidance from the Council of Europe, private entities' reasons for considering restricting content can be based on public policy reasons as well as on industrial goals.[81] While not all instances of restriction based on considerations other than public policy are necessarily incompatible with international standards on freedom of expression and human rights principles, the Act unfortunately fails to shed light on this aspect. The current provision is too vague; instead, service providers will need more guidance on what instances of content moderation and/or removal service providers are requested or allowed to take in pursuit of public policy goals, and which ones they are allowed to take in pursuit of different objectives, including those related to their own business models. Service providers should be expected to clearly indicate when content is removed in compliance with legal obligations and when measures are taken in compliance with their own terms of service.

The Act makes efforts to implement principles such as proportionality, accessibility and consistency, which is undoubtedly welcome, but at the same time robust enough mechanisms to successfully implement those principles seem to be missing. Following the recent Council of Europe's guidance, the Act should include specific and concrete provisions regarding the transparency of terms of service and reporting duties.[82]

In respect of the former point, the Act should devise a system where service providers are required to give users a clear sense of the foreseeable actions that a provider is supposed to

[81] 'Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation', pp. 35-39.
[82] Ibid, pp. 9-10, 42-45, 46.

take in respect of different kinds of content; Section 66 could be amended to include specific parameters for OFCOM to implement when issuing its guidance.

In respect of the latter, the current Act also lacks a clear metrics to assess the effectiveness of the whole system. As explained in the recent guidance from the Council of Europe, when private actors are incentivised or required by governments to remove or restrict access to content, the metrics chosen to assess the level of compliance has an implicit albeit direct impact on how service providers act in respect of different kinds of content and user behaviour.[83] For instance, it does make a difference on whether providers are incentivised to remove or restrict more or less content if their performances are going to be assessed on the basis of the speed or number of items of content removed or restricted. OFCOM should offer a detailed and specific guidance as to the information that service providers will be required to provide in compliance with their transparency obligations. Such guidance should introduce, in particular, a common framework for all regulated services to follow, and should include specific requirements regarding the format in which data needs to be provided, the level of granularity of information (e.g. the different categories of action taken in respect of different types of content, the basis for each of such actions, etc.) in order to ensure a sufficient degree of uniformity across the different entities. Otherwise, comparing and assessing results would prove difficult at best.

Age verification and end-to-end encryption

Section 12 requires service providers 'to use age verification or age estimation (or both) to prevent children of any age from encountering primary priority content that is harmful to children which the provider identifies on the service.' Service providers are required to 'prevent' and 'protect' children from accessing primary priority and priority content respectively. The same provision goes on to further explain that 'the age verification or age estimation must be of such a kind, and used in such a way, that it is highly effective at correctly determining whether or not a particular user is a child.' Similar obligations also appear elsewhere in the Act: for instance, the latest version of the Act requires providers of pornographic content to use age verification, age estimation or both.

However, age verification has been proven to be impossible to operate effectively without a user identifying themselves and also most often revealing further personal information in the process;[84] therefore, the requirement for service providers to verify users' age raises concerns for its foreseeable impact on the right to privacy and data protection. As a result, service providers will have to either utilise ineffective technologies, or resort to systems that would breach their users' fundamental rights (the right to privacy most immediately). The concern is that, for lack of viable and trustworthy technological means, service providers could simply decide to rid their platforms of content indicated in Sec. 61 and 62 for all their users – including adults.

Reflecting on such concerns, in 2021, the French authority issued an opinion[85] recommending that age verification systems should, at least, never require or allow a service provider to collect users' identity documents, or to resort to age estimation techniques based

[83] Ibid., pp. 42-45.

[84] See S van der Hof & S Ouburg, "We Take Your Word for It' - A Review of Methods of Age Verification and Parental Consent in Digital Services' (2022) 8 Eur Data Prot L Rev 61.

[85]. Commission Nationale de l'Informatique et des Libertés, 'Délibération n° 2021-069 du 3 juin 2021 portant avis sur un projet de décret relatif aux modalités de mise en œuvre des mesures visant à protéger les mineurs contre l'accès à des sites diffusant un contenu pornographique (Demande d'avis n° 21007330)' (in French), 2021.

on users' web browsing history or the processing of biometric data. Conversely, the authority recommends using a trusted independent third party for age verification. Separating the two functions (the provision of the service for which the user accepts to prove their identity, and the verification of the user's age) allows to make sure that no single entity would ever know both the identity of a user, and their browsing activities. In April 2023, the Italian data protection authority and independent media regulator started working on a joint approach to age verification, similarly considering a trusted third-party mechanism.[86]

As OFCOM starts working towards developing the relevant guidance, it would be important that it develops such guidance in collaboration with the Information Commissioner's Office (potentially going further than the mere 'consultation' currently required by the Act), to pursue the double aim of minimising the amount data collected and making the verification system reliable. The insertion of a new Clause after Sec. 143 concerning OFCOM's reports about use of age assurance, which includes considerations concerning the costs of age assurance and the need to protect users' privacy, is certainly most welcome. In a similar sense, also the new amendments to Sec. 4 requiring that OFCOM has regard to the need to strike a balance between the level of risk and the users' freedom of expression is a step in the right direction. Considering an independent trusted third-party system would be an opportunity to help minimise any negative impact on the right to privacy, OFCOM could also consider developing guidance on how service providers could better tailor their actions in respect to different age groups.

The Act also allows OFCOM to require service providers to use 'accredited technology' to detect, identify and prevent individuals from accessing terrorism and CSEA content. The wording of this provision could be interpreted as allowing OFCOM to require service providers to deploy detection tools capable of circumventing human end-to-end encryption. Recently, the Secretary of State for Science, Innovation and Technology admitted that technology enabling such controls without unduly impinging on users' privacy is still 'in development'.[87] It thus remains important that OFCOM refrains from mandating the use of over-intrusive technologies that could destructively impact on fundamental rights (the right to privacy most immediately). Collaborating with industry stakeholders and developing a regulatory framework to ensure that any technology uptake in the future is placed within a framework that is fully respectful of privacy and data protection rights should remain a priority aim.

[86] Press release, 'Garante privacy e Agcom insieme per tutelare i minori online. Istituito un tavolo di lavoro per elaborare un codice di condotta nell'ambito del protocollo d'intesa' (in Italian), 2023.

[87] BBC, 'Minister defends safety law on messaging apps', 10 August 2023.