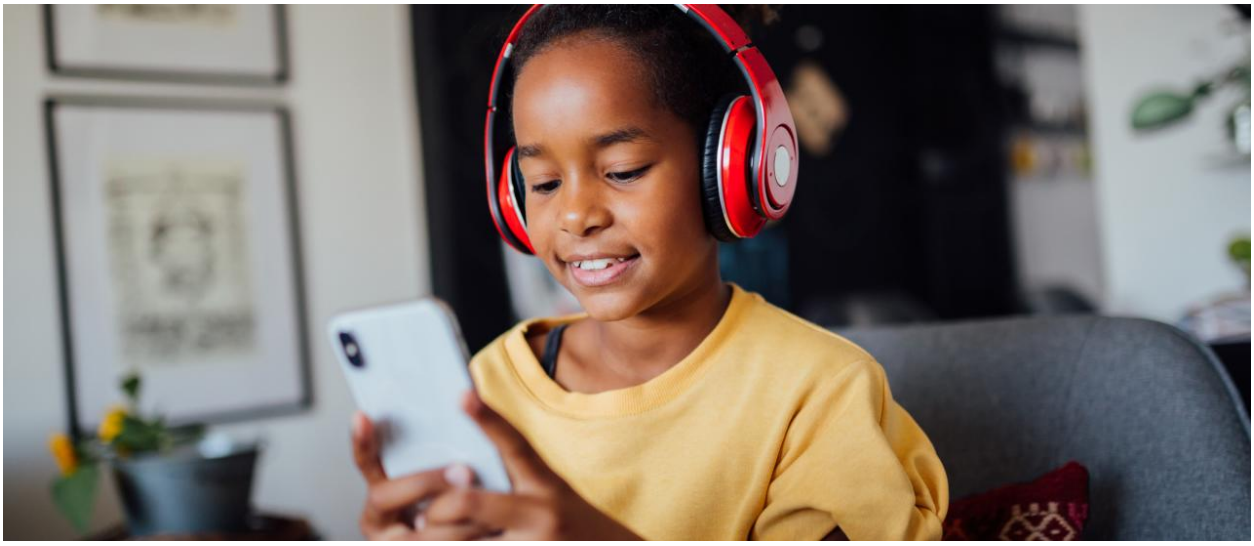
How to Create a Safe (and Open) Online Space



By François Candelon , Boston Consulting Group; Louis-Victor de Franssu (INSEAD MBA '18D), Tremau; and Theodoros Evgeniou , INSEAD

Regulatory and business challenges must be met for an internet that is both safe and allows for freedom of expression.

In a reversal of its long-held practice of "privacy first", Apple announced in August 2021 that it would launch a **new feature to scan images and videos** on its devices in order to detect stored child sexual abuse material (CSAM). The policy shift epitomises the major changes happening today both in regulations and in businesses aiming at ensuring a responsible use of technology and a safe digital space. Yet, Apple's new policy raised so many concerns from security and privacy experts that the **company has delayed its plan**.

The concept of a digital safe space is not limited to the proliferation of CSAM. Intermediary service providers, i.e. any firm that connects people, such as social media, marketplaces or online platforms for disseminating user generated content, face a growing number of abuses of their services. These include the spread of hate speech, terrorist content, illegal goods and services, spam and disinformation.

In fact, every year intermediary service providers around the world detect and remove billions of pieces of content from their platforms because the content is either illegal or contrary to their terms of service.

This affects small as well as giant platforms. Thousands of small online platforms have become home to a massive amount of illegal content posted by their users every month. Facebook identified more than 500 million pieces of such content in 2020 (1.3 billion, including spam) and **spends hundreds of millions of dollars on content moderation**. This content is so extreme and violent that people moderating it are reported to often **suffer mental health issues**.

Of course, the issue of illegal or harmful content did not appear with the rise of digital services. But the scale and speed at which such content can spread and be amplified by malicious actors who have become increasingly sophisticated, is worrying.

This has raised alarms for governments around the world which are designing new regulatory frameworks to mitigate some of these risks, with important implications not only for the future of society but also for the businesses they intend to regulate. However, achieving a safe digital space has and will continue to prove significantly challenging for regulators and companies alike.

Regulatory challenges

Democratic governments attempting to regulate the online space must grapple with contradictory objectives. They need to balance between, on the one hand, keeping the internet safe by mandating platforms to prevent the spread of illegal content and, on the other, ensuring that fundamental human rights, including freedom of speech, are protected online.

With more than **95 million photos** uploaded daily on Instagram, to name one platform giant, the sheer volume and potential for virality of content posted online makes ensuring judicial review prior to content removal nigh on impossible. Governments must therefore rely on setting out obligations for the private sector to moderate illegal content based on specific regulatory principles. But the more stringent the rules, the higher the risk of **over-content removal** and the more lenient the regulation, the higher the risks of illegal or harmful content spreading.

A related challenge for legislators is defining what effectively constitutes illegal content in a way that is broad enough to cover the targeted harms and specific enough to avoid the risks of censorship creep. Impractically broad definitions present serious risks for freedom of expression. Many worry that this difficulty could lead to political censorship in less democratic countries that would attempt to define rules without the proper safeguards.

Moreover, such regulatory definitions could leave substantial grey zones, requiring companies to decide on whether to remove content based solely on their discretion. This ambiguity combined with pressure on platforms to act as soon as such content is detected increases the risks of over-censorship, with important repercussions on freedom of expression online.

Another difficulty faced by regulators is how to implement effective obligations while ensuring competition within markets. This means finding the right balance between imposing minimum requirements for all related services without creating barriers to either innovation or market entry.

In an attempt to find fit-for-purpose solutions to these dilemmas, democratic governments and some of the largest digital services initially launched a series of self- and co-regulatory initiatives, like the **Facebook white paper** on regulation, or the **EU Code of Conduct**. Yet, outcomes were not always deemed sufficient by regulators which instead have started to develop new frameworks obliging online platforms to address detected illegal content or else face severe penalties.

In general, these new regulatory approaches can be divided into two broad categories: *content-specific* and *systemic*. The first consists of designing legislation to target a single specific type of online harm such as copyright infringements, terrorist content or CSAM and focuses on the effective and timely removal of that content. Examples of such regulations include the **European Union's Terrorist Content Online Regulation**, the **French law on disinformation**, the **German Network Enforcement Act** (NetzDG) as well as the **Directive on Copyright in the Digital Single Market**.

In contrast, the *systemic* approach aims at providing a cross-harm legal framework whereby online companies must demonstrate that their *policies, processes and systems* are designed and implemented to counter the spread of illegal content on their platforms and mitigate potential abuses of their services while protecting the rights of their users. This is the direction

proposed in the recent **Online Safety Bill** in the United Kingdom and the **Digital Services Act** (DSA) in the European Union.

In the case of the DSA for example, first presented by the Commission in December 2020, the legislators do not modify the existing liability regime, nor do they define illegal content online. Instead, the Commission sets new harmonised responsibilities and due diligence obligations for intermediary service providers: They must have in place processes and procedures to be able to either remove or disable content from their platforms when they find out that it is illegal. These regulations have implications for all intermediary service providers that go beyond potential large financial penalties.

Business challenges and implications

Firms will need to move from the culture of "move fast and break things" to a more reasonable "move fast and be responsible" as they comply with complex cross-jurisdictional demands while maintaining customers' trust. A shift towards a risk-based approach – already the path some regulators take, as the **EU proposal on regulating AI** indicates – requires organisational changes and the development of new risk management frameworks. Affected businesses need to understand the operational implications of the new regulatory obligations, assess their ability to comply and implement the appropriate risk mitigators.

Lessons from other sectors, such as finance, can prove useful. Much like in those sectors, online platforms will need to develop new policies and procedures, and then implement technical solutions. They will also need to create new roles and responsibilities, ultimately leading to organisational and cultural changes within their businesses.

New risk management processes and procedures

First, companies, regardless of their size, will need to put processes in place to address the illegal content that they have been made aware of from a number of different sources, such as national competent authorities, the platform's users or its internal moderation systems. They will also need to develop content moderation management processes and tools to ensure transparency, fairness, safety and compliance across different jurisdictions. These will unavoidably add cost and operational complexity for all online platforms.

For example, the **European Commission** estimates the annual cost of implementing and operating such tools, which includes content moderation management or **transparency reporting** workflows, can reach tens of millions annually for the **larger players**.

Second, new transparency requirements for online advertising call for online platforms to develop dedicated processes and tools to provide information to their users concerning the advertiser and their target audience. Additionally, providers of online marketplaces will also be required to enact *Know your business customer* policies and collect identification information from users operating on their platform. This obligation is largely inspired by similar requirements in the **financial industry**, adopted to limit the risks of money laundering.

And third, very large online platforms (VLOPs) will be subject to further requirements, including the obligation to conduct annual risk assessments on significant systemic risks stemming from the use of their services. These assessments will need to include risks related, for example, to the dissemination of illegal content through their services and the intentional manipulation of their platforms. While the EU Commission does not provide, at this stage, any advice on the risk assessment methodology, the DSA contains an initial list of potential risk-mitigation measures.

Organisational and cultural changes

The development of an effective risk management framework will also require the set-up of a well-balanced enterprise organisation and risk culture, aligning compliance objectives with regulatory obligations, business and growth models, and reputation risk management. In fact, through the DSA, the European Commission will require that an organisation’s chief compliance officer has sufficient financial, technological and human resources as well as the adequate level of seniority to carry out the expected tasks. While these obligations target solely VLOPs, online platforms desiring to scale and expand their business across multiple jurisdictions within the EU will benefit from early adoption of such organisational structures.

Yet organisational changes will not be sufficient by themselves. As they grow, online platforms will need to move away from a Facebook culture to one of compliance where the firm’s systemic risks are understood and where employees are empowered to do the right thing.

Fast changes

Almost two decades after the first social media platforms arrived on the internet, revolutionising the ways human beings interact, communicate and do business, we have come to a bit of an impasse. The talk about regulating these businesses has amplified globally, especially given the potential impact social media can have on our **political** and **socioeconomic** systems. These platforms can become home to different communities but also targets of illegal content postings and coordinated attacks. The upcoming regulations under development across multiple jurisdictions will not change this but will force the digital industry to adapt to a new paradigm and to find innovative solutions to tackle harmful and illegal online content.

*This is an adaptation of an **article** published in* WEF Agenda.

Find article at

https://knowledge.insead.edu/operations/how-create-safe-and-open-online-space

---

About the author(s)

François Candelon is a Managing Director and Senior Partner at the Boston Consulting Group. He is also the Global Director of the BCG Henderson Institute.

Louis-Victor de Franssu is a co-founder and CEO of Tremau.

Theodoros Evgeniou is a Professor of Decision Sciences and Technology Management at INSEAD. He has been working on machine learning and AI for over 25 years.