THE ONLINE REGULATION SERIES

–

2020

BACKGROUND TO TECH AGAINST TERRORISM

Tech Against Terrorism is a public-private partnership supported by the United Nations Counter-Terrorism Executive Directorate (UN CTED). Tech Against Terrorism was launched in April 2017 at the United Nations Headquarters in New York and is implemented by the Online Harms Foundation. As a public-private partnership, the initiative has been supported by the Global Internet Forum to Counter Terrorism (GIFCT) and the governments of Spain, Switzerland, the Republic of Korea, and Canada.

Our research shows that terrorist groups consistently exploit smaller tech platforms when disseminating propaganda. At Tech Against Terrorism, our mission is to support smaller tech companies in tackling this threat whilst respecting human rights, and to provide companies with practical tools to facilitate this process.

Our core aim at Tech Against Terrorism is to support the tech industry in building capacity to tackle the use of the internet for terrorist purposes whilst respecting human rights. We work with all types of tech companies, such as social media, pasting, file-storage, messaging, fintech platforms, and web infrastructure providers. Our core mission is providing the global tech industry with the tools needed to effectively tackle terrorist activity on their platforms..

1. **Analysis of the threat and outreach**
   We carry out extensive open-source intelligence analysis to identify platforms at risk and build constructive working relationships with the tech sector, as well as facilitating public-private cooperation

2. **Knowledge sharing and best practice**
   We facilitate intra-industry and cross-sector support mechanisms through online tools, guides, and practical datasets to support policy and content moderation decisions. Here we work closely with the GIFCT in organising global workshops and webinars. We also support companies through our membership and mentorship programmes.

   The Online Regulation Series falls within the scope of our knowledge sharing activities, as we strive to constantly provide tech companies with all the resources they need to counter terrorist use of the internet, and inscribe their efforts into the rule of law.

3. **Tech development and operational support**
   We provide technical support and resources for tech companies to improve their counterterrorism mechanisms, for example through data science or development support. Examples of past work within this workstream includes our work with Jihadology.net[1] and our current work on the Terrorist Content Analytics Platform.[2]

For more information on our organisation and how we strive to support the global tech sector and in particular smaller platforms, please visit www.techgainstterrorism.org.

[1] https://www.techagainstterrorism.org/2019/04/10/press-release-10th-april-2019-launching-an-updated-version-of-jihadology-to-limit-terrorist-exploitation-of-the-site/
[2] https://www.techagainstterrorism.org/2019/06/27/press-release-tech-against-terrorism-awarded-grant-by-the-government-of-canada-to-build-terrorist-content-analytics-platform/

BACKGROUND TO THE ONLINE REGULATION SERIES

2019-2020 witnessed many developments in terms of regulation of online speech and content, in particular in relation to countering the spread of terrorist content online. Over the past two years, several new laws have been passed or proposed to parliament in a number of countries including Australia, Brazil, France, India, the United Kingdom, Morocco, Pakistan, Singapore, Turkey, and the European Union.

Facing this fast-changing landscape, Tech Against Terrorism decided to provide smaller tech companies with a comprehensive overview of global online regulation. In doing so, we reviewed over 50 legislations, proposals, and guidelines that aim to regulate the online sphere, as well as over a 100 sources and civil society reports.

This effort culminated in the **Online Regulation Series:** For over a month, Tech Against Terrorism focused its outreach and knowledge sharing efforts on providing our stakeholders with an update on the state of global online regulation. By doing so, we hoped to have shed light on a complex, yet key issue for anyone interested in countering terrorist use of the internet whilst safeguarding human rights and freedom of expression: the regulation of online content.

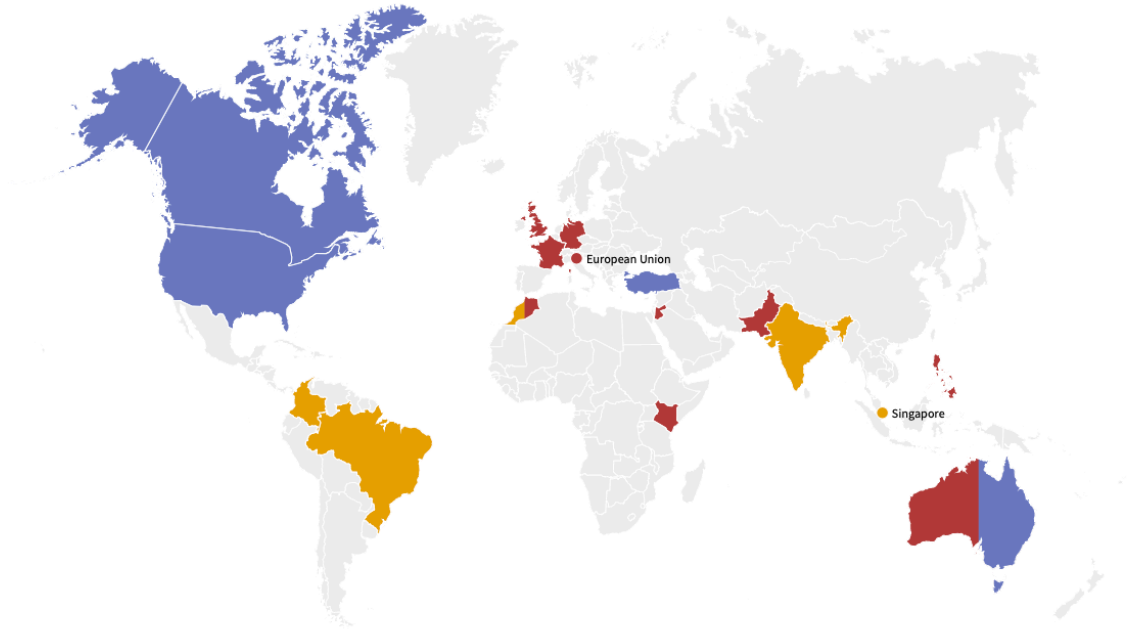
During the Online Regulation Series, we focused on three questions to improve our understanding of online regulation: What is the global state of play with regards to online regulation? What are some of the recent proposals that aims to regulate online content? What are the implications for tech platforms?

Throughout the series, we published 20 blogposts on our website**, and shared relevant resources on** Twitter**:**

- 17 country-specific blog posts divided by region: Asia-Pacific, North America, Europe, MENA and Sub-Saharan Africa, South America.
- 3 additional blogposts on tech sector initiatives and expert perspectives to complement our regional focus.

The Online Regulation Series concluded with a webinar entitled *The State of Global Online Regulation,* welcoming tech policy and digital right experts to share insights on the key global regulations that is shaping online speech around the world.

Editorial note: The analysis included in this report are as they were when published on Tech Against Terrorism's website in October – November 2020. Due to the fast-changing nature of the online regulatory landscape, some of the proposals covered have seen been passed, whilst new ones have been suggested. As the state of global online regulation continues to change, Tech Against Terrorism will strive to provide regular updates on the implications for tech companies, and their efforts in countering terrorist use of the internet whilst respecting human rights.



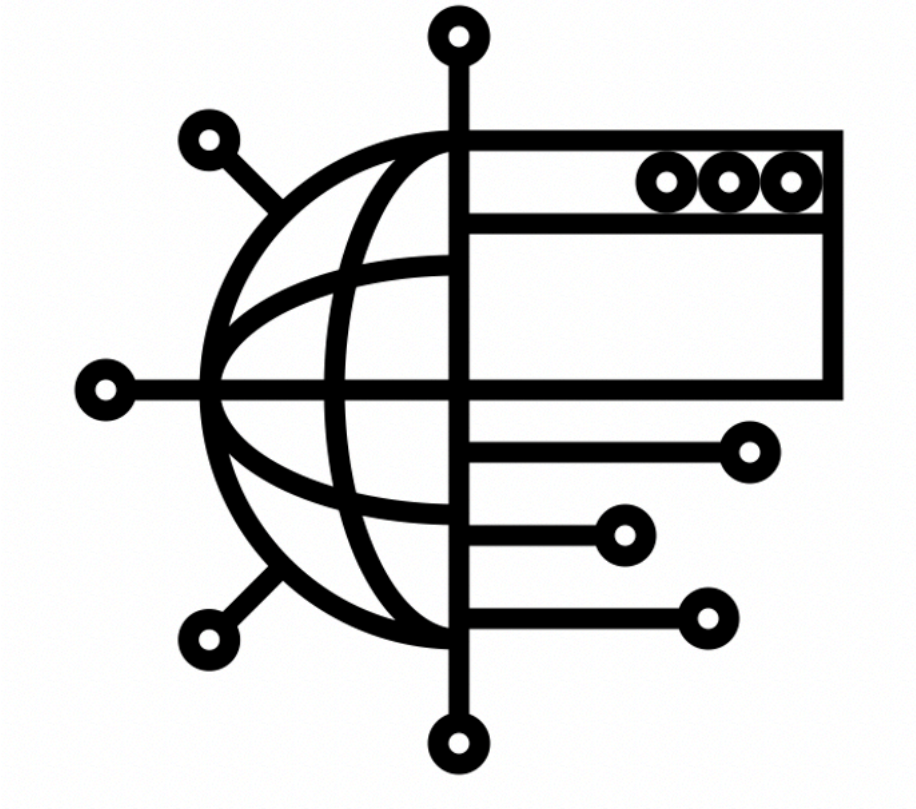
THE ONLINE REGULATION SERIES – OVERVIEW

Amongst the different trends Tech Against Terrorism has observed when conducting the Online Regulation Series, is the "rationale" behind the legislations passed or proposed:

Countering terrorist and violent extremist content, or "harmful" content: In general, these regulations target terrorist use of the internet by compelling tech companies to rap- idly remove terrorist and violent extremist content from their platforms. Often including short removal deadlines (from 1 to 24 hours) and heavy fine in case of non-compliance. The German NetzDG, 2017, was the first of such regulations and was quickly followed by others in Europe, including at the EU level. Besides terrorist use of the internet, some of these legislations also target "harmful" online content more generally. Which broadly cover anything from illegal content and incitation to hatred to suicide-promoting content, depending on the country.

Countering the spread of Misinformation and Disinformation: Countries that have faced important spread of online misinformation in recent years, notably Brazil and India, are introducing online regulations aimed at targeting this phenomenon by imposing traceability requirements for messages shared online. Or, in the case of Singapore, the possibility for the government to issue removal orders.

Adapting to the Digital Space: These legislations are motivated by the idea that existing regulations are no longer adapted to the reality and risks of today's digital world. For instance, the Digital Services Act (DSA) at the EU, has been explicitly framed as a response to how digital changes "impact our lives



SECTION 01: EXPERTS PERSPECTIVE

To complement our country-specific analysis, and offer additional insights into the state and evolution of online regulation, this second section of the Online Regulation Series offers insights into tech sector initiatives related to content governance, as well as into experts' perspective on this complex yet important topic.

THE STATE OF ONLINE REGULATION | ACADEMIC ANALYSIS

An in-depth look at academic analysis of global efforts to regulate online content and speech.

Key takeaways:

- Academics agree that global regulation of online speech has changed drastically over the past two decades, and that there has been a sharp increase in regulatory efforts over the past four years.
- Generally, academics agree that there is a need for improved regulatory measures to create a healthier online environment.
- Overall, academics are concerned that current regulatory efforts and proposals do not account for how content moderation works in practice and risk having a negative impact on freedom of expression, the rule of law, and ultimately serve tech company interests rather than producing accountability.

Background: evolution of content moderation

Academic research demonstrates that online regulation has drastically evolved since the emergence of the internet. Whilst big tech companies initially had rudimentary moderation guidelines,[3] most of them now have intricate moderation policies and mechanisms in place. Due to the fact that the most dominant global speech platforms were founded in the United States, the online speech landscape has largely been shaped by US First Amendment thinking[4]. However, academics highlight that this is rapidly changing.

Jonathan Zittrain has, in the context of analysing digital governance generally, divided the period since the emergence of the internet into three eras:

- The rights era, in which users' right to expression was prioritised by tech companies and largely accepted by the public, with objectionable content seen as a price to pay for the democratised speech culture that the internet afforded.
- The public health era, which saw companies shift towards an approach weighing the risks and benefits of allowing certain material – such as terrorist content or incitement to violence – which inevitably led to restrictions of speech on platforms.
- The process era, in which Zittrain says the digital governance field requires "new institutional relationships" that can account for the fact that not all views will or can be reconciled, but also allows for an accountable process in which such differences are settled.

[3] Both Facebook and YouTube initially had a one-page document to guide decision-making.
[4] Meaning to allow all forms of speech rather than restricting potentially harmful speech (in line with the First Amendment of the US Constitution), which many other countries do via legislation (such as Holocaust denial).

Evelyn Douek has developed on this, focusing on content moderation specifically. She describes the first era as "posts-as-trumps" where "the First Amendment's categorical and individualistic" take on speech adjudication allowed for users to "post what they wanted." Since this is no longer seen as tenable due to these policies allowing potentially harmful speech, large platforms have adopted a proportionality approach which acknowledges that free speech should be restricted in certain cases. Douek highlights that this is the dominant form of rights adjudication outside of the United States. Further, Douek argues that since content moderation is "impossible" to get perfectly right, tech companies should focus on probability. Tech companies and lawmakers alike should accept that platforms will make errors, and focus on deciding what type of errors are acceptable to produce a healthy online environment. This type of probabilistic enforcement is, according to Douek, the best solution between the extremes of "severely limiting speech or letting all the posts flow".

Platforms as de facto regulators

Academics show that regulation has, prior to the recent regulatory push, been mainly outsourced to tech companies, something which coincided with platforms taking more of a "public health" or proportionality approach to moderation. Klonick has described the larger tech companies as the "New Governors" – bodies that "sit between the state, speakers, and publishers", and are able to empower individual users and publishers".

Whilst academics disagree over the extent to which governments have spurred this trend, there is general agreement that governments, until recently, have been content to let platforms act as de facto regulators. Keller, Douek, and Danielle Citron all highlight this, noting that governments have "outsourced" policing of the internet for illegal or "harmful" content to tech platforms, something which Jack Balkin in 2014 labelled "collateral censorship." All have raised the potential downsides with what they see as a lack of accountability with this model.

Terrorist use of the internet and terrorist content has not been an exception to this rule. On the contrary, several of the mechanisms that scholars note have contributed to the "platforms as regulators" trend aim at quelling terrorist or extremist content online. Citron has highlighted the potential negative implications of this. Examining the European Union's (EU) engagement with tech platforms to tackle hate speech and extremist content, Citron argues that the EU has – via a combination of introducing voluntary industry efforts and "threats" of regulation – made tech companies become arbiters of extremist speech. According to Citron, this in turn leads to legal content being removed, something she calls "censorship creep". So-called Internet Referral Units (IRUs)[5] are often included by academics as part of this trend as well.

Academics also see some of the industry collaborative initiatives that have been created to tackle various illegal and harmful content, such as child sexual exploitation and terrorist content, as a result of government outsourcing. Douek has criticised such industry coalitions – including the Global Internet Forum to Counter Terrorism (GIFCT) – which she calls "content cartels", for their lack of accountability and transparency (more about this in our piece on tech sector initiatives).

[5] Law enforcement bodies operating within national or regional police mechanisms and reporting suspected terrorist content to tech companies for assessment and takedown against company ToS.

Government led regulation on the rise

However, as this series has shown, in recent years regulation aimed at stifling illegal or harmful online content has begun to emerge across several jurisdictions. Academics note that terrorist use of the internet, and particularly terrorist content, is at the forefront of many such regulatory efforts. Some of the landmark regulatory proposals[6] that we have covered in this series have a strong or at least partial focus on terrorist content. This is not surprising, given the seriousness of the threat. However, Daphne Keller has – in a podcast episode with us at Tech Against Terrorism – noted that there is an absence of terrorism experts in online regulation endeavours, and has warned that this leads to misguided policy proposals that risk having limited effect in terms of actually tackling terrorism and terrorist use of the internet.

It is worth examining what patterns that academics have identified across the regulation introduced in the last few years. Broadly, scholars have identified the following trends:

- Legal liability shields are being removed, made conditional, and questioned
- Removal deadlines, and fines for failing to meet them, are frequently introduced to expedite content removal
- Mandating the removal of "harmful" material, despite its legality, is increasingly included in legislation, sometimes by assessment against company Terms of Service
- Increasingly, governments are requesting that tech platforms carry out the extraterritorial enforcement of national law
- Duty-of-care models, in which regulators aim to encourage systemic change in tackling illegal and harmful speech, are increasingly investigated as options by lawmakers
- Outsourcing of adjudication on content's legality to tech companies is still pursued by governments, however now by introducing such mechanisms in law

Questioning of intermediary liability shields

The perhaps most consequential change that global regulation has touched upon is that of legal liability for tech platforms, something which they have been exempted from in the US, Europe, and various other local jurisdictions for more than two decades. Several regulations propose a move away from the current scheme under which platforms are not held legally liable for what users post on their platforms. Zittrain notes that this is not new, as intermediary liability is historically where "the most significant regulatory battles have unfolded."

There is general academic consensus that removing legal liability shields is concerning, particularly due to censorship concerns. As both Keller and Tiffany Li note, the two-decade long track record of intermediary liability laws indicate that when shields are removed, platforms will almost always err on the side of removal. However, that does not mean that academics agree that the current scheme is flawless, with some arguing that laws like Section 230 might need to change to encourage "improved" content moderation amongst tech companies (more in this in our next blogpost).

[6] Including in the European Union, the United Kingdom, France, Pakistan, and the Philippines.

Removal deadlines

Academics have noted an increase in removal deadlines in global regulation. Such deadlines compel companies to remove illegal or harmful content within a specified timeframe.[7] Failure to comply with such deadlines usually result in financial penalties. David Kaye, former UN Special Rapporteurs on Freedom of Expression, and Fionnuala Ni Aolain, the UN Special Rapporteur on Counter Terrorism and Human Rights (both of whom are academics specialising in human rights law) have warned that such short timelines will not give platforms enough time to assess content's legality, and might therefore lead to platforms removing legal content to avoid penalties.

Further, Douek has questioned the efficacy of punitive measures that focusses on individual cases (such as failure to remove content within a given timeframe). Douek argues that this will create "bad incentive problems" and will give more weight to platforms' own interests (in this case avoiding fines) rather than providing meaningful accountability. Secondly, Douek argues that removal deadlines are based on an overly optimistic belief in automated content removal tools, and that such requirements are essentially an error choice in which platforms will choose to err on the side of removal, whereas lawmakers seem to believe that platforms can remove "the bad without the good."

Mandating removal of "harmful" content

Academics have also highlighted, mostly with concern, the introduction of legislation that targets "harmful" content. The reason academics, as well as human rights activists, are concerned is the fact that "harmful" is rarely precisely defined and that several categories of potentially "harmful" speech that might be legal, and that introducing laws compelling companies to remove such content will result in the removal of legal speech.

Several academics have flagged that governments sometimes base such removal requests on company ToS. As Li notes, removing content via company Terms of Service (ToS) is often faster than going through a formal legal process. Furthermore, company ToS are often far more expansive in the "harms" they prohibit compared to national legislation. This is not surprising. As Klonick points out, companies often need to be more restrictive than national legislation out of "necessity to meet users' norms for economic viability." However, government leveraging of private companies' speech policies may have negative consequences with regards to the rule of law and accountable process. Keller has, when writing about the proposed EU regulation on online terrorist content, referred to this as the "rule of TOS", and has warned that it might lead to governments "exporting" national speech restrictions across the EU.

[7] In the proposed French law, it was 24 hours (1h for terrorist and CSA material), in the proposed EU regulation it is one hour, and in Australia companies are compelled to remove content "expeditiously" (without specifying a timeframe).

Extraterritorial enforcement of national law

Scholars note that whilst the largest tech companies have, due to their founding in the US, initially shaped their content standards on First Amendment norms, this approach has had to be adapted to match global audiences. Kate Klonick highlights how Facebook, YouTube, and Twitter all wrestled with challenges arising from their platforms allowing speech that is acceptable in American speech culture but unlawful or unacceptable in others.[8] The way companies solve this is often by "geo-blocking" content in some jurisdictions, making it invisible for users in that country whilst allowing it in other jurisdictions since it does not violate their own standards. Increasingly, governments and courts have begun to compel companies to remove access to content violating national legislation worldwide (Canada, France, Austria, and Brazil are some examples), a development which experts are concerned about due to the extraterritorial enforcement of national law.

Duty-of-care models

Some countries[9] have considered a so-called duty-of-care model. Such models aim to encourage more systemic change amongst companies as opposed to targeting illegal and harmful content via specific measures, such as removal deadlines. Many academics welcome the systemic thinking approach. Li highlights that regulation on the systemic level is likely easier and more effective than regulating content itself, particularly due to the freedom of expression concerns that such approaches entail. Similarly, Douek argues that regulation should focus on the "systemic balancing" of platforms rather than focussing on specific types of speech.

However, Keller has raised questions about the systemic duty-of-care model and how it would function alongside existing intermediary liability protections. For example, if a duty-of-care model requires companies to proactively seek out and remove content, would that mean that they are seen as active curators and therefore lose liability protections currently afforded under the EU's E-Commerce Directive or the US Section 230? Keller highlights that such a model might actually make it more difficult to hold platforms accountable, as platforms can simply point to their obligations under the duty-of-care model.

Outsourcing adjudication of illegality to the tech sector

Academics have noted that, despite the move by certain governments to regulate content more directly, several governments still rely on companies to adjudicate on content's illegality and have made this a key requirement of the law[10]. Whilst, as Douek notes, the sheer scale and technical requirements might always leave platforms as the de facto regulators of speech, there are concerns that outsourcing adjudication of content legality to private companies rather than the legal system will undermine the rule of law. According to Kaye, this lack of judicial oversight is incompatible with international human rights law.

[8] Some early encounters of this challenge being content defaming the late Thai King Bhumibol, or the founder of Turkey, Mustafa Kemal Atatürk.
[9] The most notable case being the United Kingdom.
[10] Germany's NetzDG law is one example.

Resources

Balkin (2019), *How to regulate (and not regulate) social media.*

Li (2019), *Intermediaries and private speech regulation: a transatlantic dialogue – workshop report,* Boston University School of Law.

Zittrain (2019), *Three Eras of Digital Governance.*

Caplan (2018), *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches,* Data & Society.

Citron (2018), *Extremist Speech, Compelled Conformity, and Censorship Creep*, Notre Dame Legal Review.

Klonick (2018), *The New Governors: The People, Rules, and Processes Governing Online Speech*, Harvard Legal Review.

Keller (2018), *Internet Platforms: Observations on Speech, Danger, and Money*, Hoover Institution.

Keller (2019a), *The EU's Terrorist Content Regulation: Expanding the Rule of Platform Terms of Service and Exporting Restrictions from the EU's Most Conservative Member States,* Stanford University Center for Internet and Society.

Keller, (2019b), *Who Do You Sue?,* Hoover Institution.

Keller, (2020), *Systemic Duties of Care and Intermediary Liability,* Stanford University Center for Internet and Society

Douek (2020), *Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability,* Columbia Law Review.

McDonald, Giro Correia, Watkin (2019), *Regulating terrorist content on social media: automation and the rule of law*, International Journal of Law in Context.

Kaye (2019), *Speech Police: the Global Struggle to Govern the Internet.*

THE FUTURE OF ONLINE REGULATION | EXPERTS' RECOMMENDATIONS

To follow-up on our previous blogpost on academic analysis of the state of global online regulation, we take here a future oriented approach and provide an overview of academics and experts' suggestions and analysis of what the future of online regulation might bring.

Systematic duty of care and the future of content moderation

With certain policy-makers around the world, notably in the UK, pursuing the possibility of mandating platforms to abide by a "systematic duty of care" (SDOC) for online content regulation, Daphne Keller has laid out possible models that a SDOC could follow, and their implications for tech platforms' immunity from legal liability, content moderation, human rights, and smaller tech platforms. Keller divides SDOCs into two possible models: a prescriptive one, and a flexible model.

- Prescriptive model: Under this formulation governments would set out clear rules and specify the proactive measure that platforms would be required to abide by. Thus setting a clear legal framework which could offer platforms immunity from legal liability. In practice, platforms would still have the possibility to do more than what would be required of them, "deploy[ing] novel 'Good Samaritan' efforts", meaning content moderation would not significantly change from how it is today. Except that we would witness an increase reliance on automated monitoring, such as upload filters which have long been criticised for their potential negative impacts on human rights and removing legal speech. Keller further notes that this model would have detrimental consequences for competition and innovation, as smaller platforms would have difficulties keeping up with the resources need to meet the proactive monitoring requirements.

- Flexible model: In this instance, regulators would limit their requirements to "broadly defined and open-ended obligations", which could be more adaptive to a changing and diverse landscape, but would also raise a number of questions on platforms' legal liability and whether compliance and over-compliance would grant them immunity. In general, this model would be characterised by platforms removing too much or too little depending on whether their own terms of services go beyond what would be required of them. Flexibility could also allow for more "leeway to figure out meaningful technical improvement", leading to more nuanced and diverse automated mechanisms. However, Keller stresses that in effect, this would be determined by regulators opting either for a diverse tech environment or for efficient regulation, whilst transparency would in any case be negatively impacted. Keller further predicts that if smaller tech platforms could have the possibility to deploy their own measures, it is likely that we would witness "an inevitable drift" toward SDOC being based on large platforms' practices.

Section 230: A landmark reform?

Following the Trump Administration's executive order in May 2020 directing independent rules-making agencies to consider regulations that narrow the scope of Section 230, the US witnessed a wave of proposed bills and Section 230 amendments from both government and civil society.
A 2019 report, published by the University of Chicago's Booth School of Business, suggests transforming Section 230 into a "quid pro quo benefit." Platforms would have a choice: adopt additional duties related to content moderation or forgo some or all of the protections afforded by Section 230. Paul M. Barrett embraces this concept and says lawmakers should adopt this approach for Section 230, emphasising that it provides a workable organising principle to which any number of platform obligations could be attached and that "the benefits of Section 230 should be used as leverage to pressure platforms to accept a range of new responsibilities related to content moderation". Examples of such additional platform responsibilities would include requiring platform companies "to ensure that their algorithms do not skew towards extreme and unreliable material to boost user engagement" and that platforms would disclose data on content moderation methods, advertising policies, and which content is being promoted and to whom. Barrett also calls for the creation of a specialised federal agency, or the "Digital Regulatory Agency", which would oversee and enforce the new platform responsibilities in the "quid pro quo" model, as well as would focus on making platforms more transparent and accountable.

Jack Balkin has suggested that governments make liability protections conditional, as opposed to the default, on the basis that companies "accepting obligations of due process and transparency. Similarly, Danielle Citron has argued that immunity should be conditioned on companies having "reasonable" content moderation standards in place. Such reasonableness would be determined by a judge.

Suggestions for new governance or regulation models

International human rights law

David Kaye, the former UN Special Rapporteur on Freedom of Expression, has suggested that tech companies ground their content moderation policies in international human rights law (IHRL). Kaye argues that this is the best solution to solve several of the challenges highlighted by academics in our previous post. For example, international human rights law offer a global structure (as opposed to national law), and provide a framework for ensuring that both companies and governments are complying with human rights standards in a transparent and accountable manner. Further, Kaye notes that Article 19 of the International Covenant on Civil and Political Rights (ICCPR) – which mandates freedom of expression – also provides for cases where speech can be restricted, where necessary to protect others' rights, and where necessary for public health and national security. Kaye argues that this means that platforms will be able to take action on legitimately harmful and illegal content.

Evelyn Douek, has whilst acknowledging that this approach has several benefits, questioned whether it will be efficient. Douek notes that there is a "large degree of indeterminacy" in IHRL, which according to her means that it will be up to platforms to assess content against such standards. Further, Douek worries that such standards could in theory provide companies with a basis for

allowing legitimately harmful content to remain online (or vice versa), since platforms and local speech culture might differ in their interpretation of the IHRL.

Social media councils

Civil society group Article 19 has suggested the creation of an independent "Social Media Council". They argued that this would increase accountability and transparency with regard to content moderation, without government restricting on speech via regulation targeting online content. The Council would be based on a "self-regulatory and multi-stakeholder approach" with "broad representation" from various sectors, and would apply human rights standards in content moderation review. Loosely based on other self-regulatory measures such as press regulatory bodies, the Council would not be legally binding but participating platforms would commit to executing council decisions.

This suggestion was supported by David Kaye and the Stanford University's Global Digital Policy Incubator (GDPi). Following a working meeting discussing the suggestion, GDPi proposed that the social media council should avoid adjudicating specific cases and instead develop and set core guidelines for companies. Article 19 differed, advocating for the Council to have an adjudicatory role and serve as an appeals and review body, with a first version being launched on a national scale as a trial.

Resources

Keller (2020a), *Systemic Duties Of Care And Intermediary Liability, The Center for Internet and Society,* Stanford University.

Keller (2020), *Broad Consequences Of A Systemic Duty Of Care For Platforms, The Center for Internet and Society,* Stanford University.

Citron and Wittes (2017), *The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity*, Fordham Law Review.

Citron and Franks (2020a), *The Internet As a Speech Machine and Other Myths Confounding Section 230 Reform,* Boston Univ. School of Law, Public Law Research Paper.

Citron (2020b), *Section 230's Challenge to Civil Rights and Civil Liberties,* Boston Univ. School of Law, Public Law Research Paper.

Article19 and UN Special Rapporteur on Freedom of Opinion and Expression (2019), *Social Media Councils - from concept to reality.*

McKelvey, Tworek, and Tenove (2019*), How a standards council could help curb harmful content online*, Policy Options.

Balkin (2020), *How to Regulate (and not regulate) social media*, Yale Law School, Public Law Research Paper.

Douek (2020), *The Limits of International Law in Content Moderation*, UCI Journal of International, Transnational, and Comparative Law (forthcoming 2021).

Barrett (2020a), *Regulating Social Media: The Fight Over Section 230 — and Beyond,* NYU Stern.

Barrett (2020b), *Why the Most Controversial US Internet Law is Worth Saving*, MIT Technology Review.

EXPERTS PERSPECTIVE | TECH SECTOR INITIATIVES

Although regulation frameworks of terrorist and harmful content online have been passed by governments in recent years, regulation in practice remains mostly a matter of solo or self-regulation by the tech sector. That is, when companies draft and apply their own rules for moderating user-generated content on their platforms or when they voluntarily comply with standards shared amongst the tech sector (the Global Internet Forum to Counter Terrorism is one example), without such standards being enforced by law.

This, coupled with increased public pressure to address the potential harmful impact of certain online content – in particular terrorist material – has led major tech companies to develop their own councils, consortiums, and boards to oversee their content moderation and its impact on freedom of speech online. In this blogpost, we provide an overview of some of the prominent tech sector initiatives in this area.

Key takeaways:

- Major tech platforms are creating ambitious oversight and advisory bodies to address concerns about their content moderation policies and practices.
- Such bodies aim to increase accountability and transparency by, for example:
    - Providing an extra instance for user appeals.
    - Providing insight into a platforms' practical decision-making in the content moderation process.
    - Providing external expert guidance on policies.
- Collaborative industry efforts such as the Global Internet Forum to Counter Terrorism (GIFCT) aim to provide practical capacity building and knowledge sharing for tech companies, and have also launched their own research network.

Global Internet Forum to Counter Terrorism (GIFCT):

The GIFCT was founded in 2017 by Facebook, Microsoft, Twitter and YouTube to facilitate collaboration and knowledge sharing amongst the tech sector to tackle terrorist use of the internet. Since its founding, the GIFCT, which runs its own membership programme, has grown to a dozen members and has taken a prominent role in the Christchurch Call to Action – launched following the March 2019 attack in Christchurch, New Zealand, which was livestreamed on Facebook.

Tech Against Terrorism has been a core partner to the GIFCT since its inception, organising its inaugural workshop in San Francisco in 2017. Since then, Tech Against Terrorism has been running the GIFCT knowledge sharing programme by organising workshops and e-learning webinars, as well as implementing a mentorship programme to assist companies in meeting GIFCT's membership requirements.

In 2019 the GIFCT announced that it would become an independent organisation. This was formalised in 2020 with the hiring of its first Executive Director, Nicholas Rasmussen. The foundational goals of the new organisation include empowering the tech sector to respond to terrorist exploitation, enabling "multi-stakeholder engagement around terrorist and violent extremist misuse of the Internet", promoting dialogue with civil society, and advancing understanding of the terrorist and violent extremist landscape "including the intersection of online and offline activities."

The independent GIFCT's structure is complemented by an Independent Advisory Council (IAC) made up of 21 members representing the governmental (including intergovernmental organisations) and civil society sectors, and covers a broad range of expertise related to the GIFCT's areas of work, such as counterterrorism, digital rights, and human rights. The IAC is chaired by a non-governmental representative, a role currently held by Bjorn Ihler, a counter radicalisation expert and founder of the Khalifa-Ihler Institute. The four founding companies are also represented via the Operating Board, which appoints the Executive Director and provides the GIFCT's operational budget. Other members of the board include one other member company (on a rotating basis), a rotating chair from the IAC, and of new members that meet "leadership criteria".

The GIFCT also runs the Hash-Sharing Consortium to help member companies moderate terrorist content on their platforms. The consortium is a database of hashed terrorist content.[1] Members can add hashes of content they have previously identified to be terrorist material on their platforms to the database. All companies using it are able to automatically detect terrorist material on their platforms and prevent its upload. The Consortium was set up by the four founding companies in 2016.

Whilst the GIFCT states that "each consortium member can decide how they would like to use the database based on their own user terms of service", critics have raised concerns over the lack of transparency surrounding the use of the database and the removal of content it contributes to. However, the GIFCT has to date published two transparency reports, which provide insights into the hash-sharing database and the type of content that was added to it.[2] In the 2020 report, the GIFCT said that the hash-sharing database contained content across the following categories:

- Imminent Credible Threat: 0.1%
- Graphic Violence Against Defenseless People: 16.9%
- Glorification of Terrorist Acts: 72%
- Radicalization, Recruitment, Instruction: 2.1%
- Christchurch, New Zealand, attack and Content Incident Protocols (Christchurch, 6.8% Halle attack, 2% Glendale attack 0.1%)

Academic and online regulation expert, Evelyn Douek, has used the GIFCT as an example when cautioning against the role played by industry initiatives aiming to curb harmful online content, a phenomenon she calls "content cartels". In her analysis, Douek stresses what she sees as risks of collaborative industry arrangements including both larger and smaller companies, where "already powerful actors" can gain further power as they are able to set content regulation standards for the smaller platforms. In particular, she argues that such arrangements leave little room for challenging

the standards they set – including, in some cases, what they consider to be terrorist or harmful content.

Facebook Oversight Board

Facebook announced in 2018 that it would set up an independent "Supreme Court" to decide on complex content moderation issues for user-generated content on both Facebook and Instagram. The Facebook Oversight Board was announced a year later, in September 2019, and its first members in 2020. The Board began accepting cases in October 2020.

The goal of the Board is to "protect free expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies." The board is set up as a last appeal instance for users who wish to contest the removal of their content, and whose appeal has already been rejected twice by Facebook internal appeal process. For now, the Board will limit its oversight to content that has already been removed from Facebook or Instagram. However, Facebook has stated that the scope of the Board will be expanded to allow users to appeal for content they want to be removed from the platforms. In selecting and handling cases, the Board will focus on cases that have significant impact on online freedom of expression and public discourse, real-world impact, or "raise questions about current Facebook policies". Facebook itself can submit "urgent [cases] with real-world consequences" for review.

Besides advising Facebook on whether to allow or remove content, the Board can also "uphold or reverse a designation that led to an enforcement", such as a designation leading to the removal of a page on the grounds of terrorism. Board decisions will function as caselaw and will help influence Facebook's content moderation policies. Beside this, the Board will be able to provide direct policy guidance to Facebook on its policies and processes.

Whilst the concept of the Oversight Board has been welcomed, it has nonetheless drawn criticisms. One concern relates to the fact that the Board's charter: "still provides Facebook some leeway about how to implement the board's decisions. Critically, it only has to apply the decision to the specific case reviewed, and it's at the company's discretion to turn that into blanket policy". In particular, Facebook has stated that it would "support the Board" depending on whether implementing a decision to other cases or as policy guidance is "technically operationally feasible", and on the resources it would take the company to do so.

Kate Klonick – an expert on online speech governance – has summarised the different reactions and criticisms addressed to the Board. Amongst the main criticisms are concerns over how the Board could negatively impact Facebook's content moderation by encouraging it to either under-moderate or over-moderate; that the Board is, effectively, a PR stunt; or that it risks not being scalable. Klonick commented on these concerns by underlining the Board's potential to have a broader impact on Facebook policies, beside single cases, and on how it "might lead to more widespread user participation in deciding how to design private systems that govern our basic human rights."

Concerned with the fact that the Board would not be up-and-running by the time of the US elections, a "group of about 25 experts from academia, civil rights, politics and journalism" led by the UK-based

advocacy group The Citizens, set up their own "Real Facebook Oversight Board" in September 2020. The group set out to organise weekly public meetings on Zoom to scrutinise a broad range of issues linked to Facebook's moderation practices. Commenting on this initiative, Klonick described it as "misleading", given that it would not hear any user appeals.

Twitch Safety Advisory Council

Twitch, the leading global live-streaming platform, announced the creation of its Safety Advisory Council in May 2020. The Council's mission is to advise Twitch in its decision-making process and policy development. This includes drafting new policies, helping developing product and features for moderation, as well as promoting diversity and the interests of marginalised groups on the platform.

The Council is made up of 8 members representing a mix of Twitch creators, experts in online safety (including cyberbullying), and in content moderation. The mix of experts and creators is meant to ensure that the Council has "a deep understanding of Twitch, its content and its community". Amongst the experts is Emma Llanso, Director of the Free Expression Project at the Center for Democracy & Technology, and an expert on free expression online and intermediary liability (Emma has previously guested our podcast and our webinar series).

TikTok's Content Advisory Council

Video-sharing app TikTok unveiled its Content Advisory Council in March 2020. In a drive to improve its accountability and transparency, TikTok also announced its Transparency and Accountability Center, and has proposed the creation of a Global Coalition to Counter Harmful Content.

The Coalition is meant to target the challenges posed by the constant posting and re-posting of harmful content that all tech platforms face, and to do so via collaborative efforts between tech platforms and the "development of a Memorandum of Understanding (MOU) that will allow us to quickly notify one another of such content."

The Council, for its part, is made up of several tech and safety experts, and will advise TikTok around its content policies and practices. TikTok has announced that the Council would meet regularly with its US leaders "to discuss areas of importance to the company and our users", such as the platform integrity and policies related to misinformation.

The Council is chaired by Dawn Nunziato, an expert on free speech and content regulation at George Washington University, and includes different tech policy, online safety, and young mental health experts, with the plan to grow to about 12 experts.

Resources

Article19 (2019), *Social Media Councils: Consultation.*

Bijan (2020), *Twitch establishes a safety advisory council to help it sort out its rules,* The Verge.

Botero-Marino, Greene, McConnell and Thorning-Schmidt (2020), *We Are a New Board Overseeing Facebook. Here's What We'll Decide,* The New York Times.

Constine (2018), *Facebook will pass off content policy appeals to a new independent oversight body*, TechCrunch.

Constine (2019), *Facebook's new policy Supreme Court could override Zuckerberg,* TechCrunch.

Douek (2020), *The rise of content cartels*, Knight 1rst Amendment Institute.

Ghaffary (2020), *Facebook's independent oversight board is finally up and running,* Vox.

Ghosh (2019), *Facebook's Oversight Board Is Not Enough,* Harvard Business Review.

Harris (2019), *Establishing Structure and Governance for an Independent Oversight Board,* Facebook News Room.

Klonick (2020), *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, The Yale Law Journal.

Perez (2020a), *TikTok brings in outside experts to help it craft moderation and content policies,* TechCrunch.

Perez Sarah (2020b), Twitch announces a new Safety Advisory Council to guide it is decision-making, TechCrunch.

Radsch (2020), *GIFCT: Possibly the Most Important Acronym You've Never Heard Of,* JustSecurity.

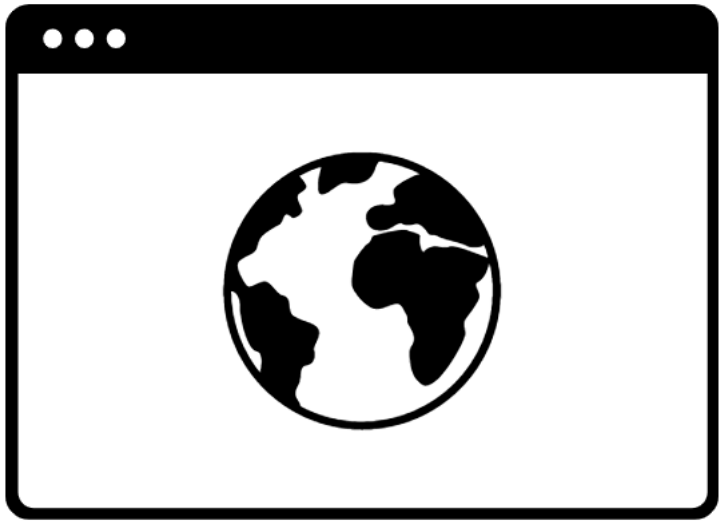
Reichert (2020), *TikTok now has a content advisory panel,* CNET

Solon (2020a), *While Facebook works to create an oversight board, industry experts formed their own,* NBC News.

Solon (2020b), *Months before it starts, Facebook's oversight board is already under fire,* NBC News.

Windwehr and York (2020), *One Database To Rule Them All*, Vox-Pol.

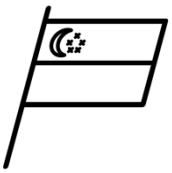
Zuckerberg (2019), *Facebook's Commitment to the Oversight Board.*



SECTION 01: GLOBAL ONLINE REGULATION

To provide a comprehensive and global overview of the different legislations and proposals aimed at regulating online content and speech, we took a country-specific approach to the Online Regulation Series. Dividing our work by geographical regions, this section offers an analysis of the state of online regulation in 17 countries across the globe.

ASIA-PACIFIC | SINGAPORE



Singapore is often deemed to be Asia's main tech hub and a top global alternative to the Silicon Valley. Many of the world's major tech platforms – including GIFCT founders Facebook, Microsoft, Google and Youtube – have their headquarters for the Asia Pacific region in Singapore. The government has been active in supporting the tech sector, advocating for an approach that promotes industry self-regulation and strong intellectual property laws.

Singapore's regulatory framework:

- *Internet Code of Practice*, October 2016, which sets baseline obligations for Internet services and content providers operating in Singapore.
- *Internet Regulatory Framework,* which provides an overview of the country's approach to online regulation and links to the *Code of Practice*
- *The Protection of Online Falsehoods and Manipulation Bill (POFMA), October 2019,* which addresses the spread of misinformation through correction and removal orders.

Main bodies overseeing online regulation:

- Infoccom Media Development Authority (IMDA), the government agency regulating the "infocomm and media" sectors in Singapore

Key takeaways for tech platforms:

- Singapore's regulatory framework does not specifically target online terrorist content. However, the prohibition of online content that incite or endorse hatred and strife can be used as a justification to remove terrorist material.
- All internet content and service providers operating In Singapore need to comply with the *Internet Code of Practice*, which effectively provides a legal basis for the prohibition of "objectionable" material.
- If in violation, the Media Development Authority "has the power to impose sanctions, including fines"[11] on tech companies.
- Under the POFMA, Government ministers can order individuals and online platforms to post corrections or take down content that is assessed by the minister to be false or "against the public interest".
- Tech platforms that do not comply with a correction or removal order under POFMA face penalties "up to S$1,000,000 per day for every day the content remains uncorrected/unremoved."

[11] Internet Code of Practice, Infocomm – Media Development Authority, Singapore, 2016

Singapore's Internet Regulatory Framework

Branding itself as "one of the most-connected countries in the world", Singapore has developed an *Internet Regulatory Framework* that sets the scene for its approach to online regulation. In general, online regulation in Singapore is considered through the lens of media development, falling under the responsibility of the IMDA. Whilst promoting industry self-regulation, media literacy, and cyber wellness, Singapore's regulatory framework has at its core the idea of "national cohesion and public interest".[12] For this reason, the existing framework focuses on content that can be deemed "objectionable" – "harmful to Singapore's racial and religious harmony, or against national interest."

In light of this, the main requirements for tech platforms operating in Singapore are laid out in the *Internet Code of Practice.* Most importantly, it covers what online material is prohibited in the country. Here again, the idea of public interest and security, as well as of national harmony, are prevalent.

The idea of ensuring that online content does not hurt Singapore's "national harmony" and "public interest" is also reflected in the Content Regulation Guidelines set out by IMDA – which also oversees the *Regulatory Framework* and *Code of Practice.* Indeed, whilst the guidelines encourage the importance of co-regulation with the industry,[13] they also have at their core the idea of "reflecting societal values and community standards", and encouraging platform to be "socially responsible" in ensuring that content "meets with community standards".

A legal framework for prohibiting incitement

Whilst the *Framework* states that it does not engage in the monitoring or restriction of individual's access to online content, the *Code of Practice* does set some baseline prohibitions on the use of the internet. Even though regulations in Singapore do not specifically target the issue of terrorist use of the Internet, certain prohibitions laid out in the *Code of Practice* provide a legal baseline for the moderation of certain terrorist online material. Mainly, the prohibition of online content that "glorifies, incites or endorses ethnic, racial or religious hatred, strife or intolerance" can be applied to content used for terrorist incitement.

So-called fake-news bill

In May 2019, Singapore introduced a new bill related to the regulation of online content: *The Protection of Online Falsehoods and Manipulation Bill* (POFMA) – came into effect in October 2019. The bill is meant to address the (viral) propagation of false statement on the internet, especially those deemed to be against the public interest by the government. Under this new legislation, any government minister "will have the power to direct individuals, publishers, internet platforms, and

[12] Singapore's Internet Code of Practice underlines the government's emphasis on ensuring societal cohesion in the City-State, especially what it labels the country's "racial and religious harmony". Following the Christchurch attack, the government of Singapore particularly acknowledged the threat that the combination of violent extremism, terrorism, and internet technology, to the country's "social harmony". On the importance of "social harmony" in the country, see: Ms Grace Fu, Minister for Culture, Community and Youth, "Preserving Singapore's social harmony in the face of emerging threats", key note address at the at the Roses of Peace Youth Forum *Aftermath of Christchurch – Lessons for Singapore,* Ministry of Culture, Community and Youth, 30 March 2019

[13] Infocomm also encourages member of the public to signal "objectionable content" to the IMDA.

mainstream media to either post corrections or takedown content, based on what the minister deems to be content deemed false and 'against the public interest'."

Civil society groups and digital rights experts have expressed concerns regarding POFMA's potential to negatively impact freedom of expression in the country and hinder public debate. Most concerns revolve around the expansive scope of the law, especially around what is "public interest", and the fact that government ministers are to make decisions on what constitute misinformation and false content. These concerns were raised by David Kaye, UN Special Rapporteur on Freedom of Expression, in a letter to the government of Singapore. Kaye also stressed that POFMA risks pushing tech platforms to over-restrict content, including lawful material.

Following the enactment of POFMA, major tech platforms – including Facebook, Twitter and Google Search – were granted a temporary exemption, which was later suspended due to the Covid-19 crisis. Whilst this exemption acknowledges that tech platforms needs an adaptation period to a legislation that can significantly impact how they moderate content, it also raises questions regarding the non-consideration of smaller tech platforms which might face more difficulties to implement the POFMA bill – especially due to a lack of technical and human resources. Given the penalties that companies can face when not complying, this bears the risk of reduced competition in the tech sector if smaller platforms are not able to catch up or are financially afflicted by the fines.

Moderation of encrypted platforms

The POFMA bill also covers the question of moderation and regulation of closed platforms, including those based on end-to-end-encryption (E2EE), such as messaging apps like WhatsApp, Signal, Wire, or Telegram. Indeed, the so-called "fake news" law does not only apply to social media and content-hosting platforms, but also encrypted apps – however the regulation of encrypted platforms under POFMA has not been enforced so far. The debate surrounding the regulation of E2EE is not specific to Singapore, and many countries have in recent years expressed their desire to regulate E2EE, often citing counterterrorism and counter child sexual abuse reasons (including the UK and the US). However, Singapore is the first to have turned this into law, therefore setting a precedent for regulation of E2EE platforms. Tech companies have fought back against such regulations on the grounds that they risk undermining the security and privacy of users, stressing the technical difficulties of creating a "government backdoor" for security reasons that would not expose users to malevolent actors.

Resources

Amnesty International (2020), *Singapore: Social media companies forced to cooperate with abusive fake news law.*

Asia Internet Coalition (2020), *Toolkit – Addressing online misinformation through legislation.*

Asia One, *A look into Twitter's Asia-Pacific headquarters in Singapore.*

Chen and Chia (2019), *Singapore's latest efforts at regulating online hate speech*, Singapore Management University, Research Collection School of Law.

Chong (2018), *Facebook's Asia team moves to gigantic new headquarters in Singapore,* CNet.

Consultancy.org (2020), *Singapore considered top alternative tech hub to Silicon Valley.*

EDB Singapore (2019), *Tech firms head to Singapore amidst Southeast Asia's growth.*

Grace Fu, Minister for Culture, Community and Youth, (2019), *Preserving Singapore's social harmony in the face of emerging threats, key note address at the at the Roses of Peace Youth Forum Aftermath of Christchurch – Lessons for Singapore.*

Infocomm – Media Development Authority (2016), *Internet Code of Practice,* Government of Singapore

Infocomm – Media Development Authority, *Internet Regulatory Framework*, Government of Singapore

Infocomm – Media Development Authority, *Internet*, Government of Singapore.

Republic of Singapore, Government Gazette, Acts Supplement (2019), *Protection from Online Falsehoods and Manipulation Act.*

Kaye, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2019), *Letter to the Government of Singapore on The Protection of Online Falsehoods and Manipulation Bill,* Office of the united Nations High Commission for Human Rights

Mullin (2019), *Fancy New Terms, Same Old Backdoors: The Encryption Debate in 2019*, Electronic Frontier Foundation.

Newman (2020), The EARN IT Act Is a Sneak Attack on Encryption, Wired.

Pfefferkorn (2020), *THE EARN IT ACT: how to ban end-to-end encryption without actually banning it?*, The Center for Internet and Society.

Republic Of Singapore, Government Gazette, Acts Supplement, *The Protection of Online Falsehoods and Manipulation Bill,* Government of Singapore.

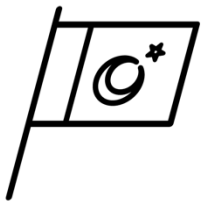
Sabbagh Dann (2020), *MI5 chief asks tech firms for 'exceptional access' to encrypted message*", The Guardian.

Ungku (2019), *Facebook, rights groups hit out at Singapore's fake news bill*, Reuters.

United Nations Office on Drugs and Crime (2012), *Use of the Internet for terrorist purposes.*

Wong (2019), *US, UK and Australia urge Facebook to create backdoor access to encrypted messages,* The Guardian.

ASIA-PACIFIC | PAKISTAN



Over the last five years, Pakistan has introduced various measures aimed at regulating terrorist content online, including the 2020 *Citizen Protection (Against Online Harm) Rules* which directly targets content posted on social media, and the 2016 *Prevention of Electronic Crimes Act* which prohibits use of the internet for terrorist purposes.

These regulations supplement the Anti-Terrorism Act of 1997 (ATA) that provides the baseline legal framework for counterterrorism measures in the Country. The ATA did not specifically target terrorist use of the internet, however, it does consider the dissemination of digital content "which glorifies terrorists or terrorist activities" to be an offence – under section 11W. The same section also prohibits the dissemination of content that incite to hatred or "gives projection" to a terrorist actor.

Pakistan's regulatory framework:

- *Anti-terrorism Act,* August 1997, which sets the framework for Pakistan's counterterrorism response
- *Prevention of Electronic Crimes Act* (PECA), August 2016, which provides a "comprehensive legal framework" to counter electronic crimes and related investigations
- *Citizen Protection (Against Online Harm) Rules,* January 2020, which regulates social media content, including terrorist material and hate speech. Following criticisms, the rules are currently being reviewed by the government.

Main bodies overseeing online regulation:

- Pakistan Telecommunication Agency (PTA), which oversees the PECA
- National Coordinator, which oversees the implementation of the 2020 Rules, appointed by the Ministry of Information and Technology

Key takeaways for tech platforms:

- Via the PECA and *2020 Citizen Protection Rules*, Pakistan explicitly prohibits terrorist use of the internet, and terrorist content shared on social media
- Under the *2020 Rules, S*ocial media platforms would:
  - be asked to remove or block access to unlawful content when notified by the relevant authorities
  - have to register and have a physical office in Pakistan to operate in the country
  - establish database center in the country within 12 months, to record and store data and online content
- Companies that fail to abide by the *2020 Rules* can be blocked from operating in the country or face detrimental fines
- Under the PECA, individuals posting terrorist material online can also be held liable and face jail terms

2020: an attempt a directly regulating social media content

In 2020, the *Citizen Protection (Against Online Harm) Rules,* specifically targeting content hosted on social media platforms was enacted. These Rules aims at curbing harmful online content, including terrorist and extremist content, as well as hate speech and misinformation.

The *Rules* require social media companies to remove or block access to "unlawful" material when requested to do so by a newly created "National Coordinator" authority. Social media companies will thus be obliged to respond to a request by, the PTA or the National Coordinator, to remove material they deem "unlawful" within 24 hours, or six hours in emergency cases. They will have three months to register with authorities in Pakistan and must have a physical presence in the country. When required to so, the companies will have to provide subscriber information, traffic data, content data and any other information or data that is sought, the rules stipulate. Companies that do not comply with the new regulation risk being blocked online and face a fine of over 3,000,000 US dollars.

This new regulation has drawn criticism from civil society groups due to what they argue are risks regarding freedom of expression. Media Matters for Democracy, a Pakistani non-governmental organisation, called the rules a "direct threat to Pakistan's digital economy and the citizens' rights to freedom of expression and privacy". Article19 has also raised concerned regarding the requirement to deploy "'proactive measures' to ensure the prevention of livestreaming on their platforms of any content in breach of any law or rules in force in Pakistan", which risks amounting to "prior censorship". Under this requirement, terrorist and extremist content being mentioned as content of "special concern" without being defined in reference to any existing legislation in the country. Article19 is also concerned with such proactive measures being synonymous to automated filter, which risks the removal of lawful content.

Criticisms at the *Rules* – including a letter from the Asia Internet Coalition reporting that there is a risk of major tech companies pulling out of Pakistan if the rules were enforced – led the government to suspend and reconsider the rules, conducting an "extensive and broad-based consultation process with civil society and technology companies."

A petition has also been filed before the Islamabad Hight Court, challenging the *Rules* on the ground that it not the prerogative of the federal government to frames rules under the PECA for removal or blocking of online content. The petition was scheduled to be hear by the court in August.[14]

A direct and comprehensive prohibition of terrorist use of the internet

These rules were promulgated under *the Prevention of Electronic Crimes Act.* Passed on 11 August 2016, this legislation is meant to provide a regulatory framework for "the prevention of electronic crimes, more specifically to "prevent unauthorized acts with respect to information systems and provide for related offences as well as mechanisms for their investigation, prosecution, trial and international cooperation". Beyond electronic crimes, the bill has been labelled as a "flagship" legislation for Pakistan counterterrorism efforts, for its direct prohibition of terrorist use of the internet.

[14] At the time of writing, there is still a lack of clarity on the status and enforcement of the *Rules.*

Indeed, the act prohibits the use of information systems for the glorification of terrorism, "recruitment, funding, and planning of terrorism", as well as cyberterrorism and hate speech. Depending on the offence, individuals can be sentenced to up to 7 or 14 years (for cyberterrorism) of prison, or a fine of over 70,000 euros.

However, the PECA has been criticized by human rights activists and civil society groups for what they deem to be an overly broad language on what constitutes illegal content, and for granting regulators, the Pakistan Telecommunication Agency, a certain leeway in deciding what is to be considered illegal content. According to critics, this bears the risks of users self-censoring their online speech. As well as the possibility to block private This has led certain human rights activists to criticise the law for the risk it poses to freedom of expression and users' privacy.

Civil society groups have also raised concerns regarding the broad jurisdiction the PECA aims to cover, as it also targets electronic crimes committed by Pakistani nationals outside of the country. An international aspect of the bill that, according to the Electronic Frontier Foundation, "could have practical consequences for the thousands of overseas Pakistanis working in the IT and infosecurity industries, as well for those in the Pakistan diaspora who wish to publicly critique Pakistani policies."

Resources

Pakistan Government (1997), *Anti-Terrorism Act.*

Pakistan Government (2016), *Prevention of Electronic Crimes Act.*

Pakistan Government (2020), *Citizen Protection (Against Online Harm) Rules.*

Article19 (2016), *Pakistan: An Analysis of the Prevention of Electronic Crimes Bill 2015.*

Article19 (2020), *Pakistan: Online Harms Rules violate freedom of expression.*

Asia Internet Coalition (2020), *AIC Submits Response to Pakistan's Citizens Protection Rules (Against Online Harm).*

Committee to Protect Journalists (2020), *Pakistan government secretly passes strict social media regulations.*

CPU Media Trust (2020) *Pakistan government secretly passes strict social media regulations.*

Digital Rights Foundation (2020), *Citizens Protection (Against Online Harm) Rules, 2020: Legal Analysis.*

Digital Rights Foundation (2020b), *DRF Condemns Citizen's Protection (Against Online Harm) Rules 2020 as an Affront on Online Freedoms.*

Farrell (2020), *How Big Tech defeated Pakistan's censorship police.*

Garg (2020), *Pakistan's Online Harm Rules: Rights to Privacy and Speech Denied",* Jurist.org.

Global Network Initiative (2020), *GNI Expresses Serious Concern Regarding Pakistan's Rules Against Online Harm.*

Kaye David, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2015), *UN expert urges Pakistan to ensure protection of freedom of expression in draft Cybercrime Bill,* UN OHCHR.

Kamran (2020), *Civil society bodies declare the 'rules for protection against online harm' a political move to silence critics; demand immediate de-notification*, Digital Rights Monitor.

O'Brien (2016), *The Global Ambitions of Pakistan's New Cyber-Crime Act,* Electronic Frontier Foundation.

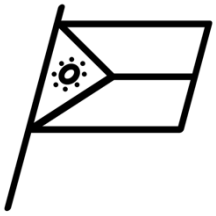
Pakistan Today (2020), *Cabinet approves new rules to regulate social media.*

Phoneworld (2020), *Government to Review Citizen Protection (Against Online Harm) Rules, 2020.*

Reuters (2016), *Pakistan passes controversial cyber-crime law.*

Shahzad (2020), *Pakistan's government approves new social media rules, opponents cry foul*, Reuters.

ASIA-PACIFIC | THE PHILIPPINES



The Philippines is one of the countries worst affected by terrorism in the world, ranking as the ninth most affected country in the 2019 Global Terrorist Index. The country has long been investing in its counterterrorism apparatus and there have been some signs that the Philippines might introduce legislation that targets online terrorist content. This is to be understood in the context of a growing internet penetration rate and increased use of social media (+8.6% in 2019-2020), coupled with growing concerns for how terrorists use the internet in the country.

Philippines' regulatory framework:

- *Anti-Terrorism Act* (ATA), July 2020, providing the legal framework for the country's counterterrorism response.
- *Cybercrime Prevention Act* (CPA), September 2012, the country's regulatory framework for the use of information and communication technologies.

Key takeaways for tech platforms:

- To date, tech companies have been exempted from liability for user-generated content posted on their platforms.
- However, recent suggestions to expand the Anti-terrorism Act of 2020 to allow for the regulation of social media indicate that making tech platforms liable for online terrorist content is not absent from the public debate in the Philippines.

Counterterrorism legislation and incitement to terrorism

Whilst the South-East Asian country does not directly regulate online content, the recent Anti-Terrorism Act provides a basis for the criminalisation of terrorist content online via its penalisation of terrorist incitement. The ATA was signed into law in July 2020, replacing Philippines' previous counterterrorism legislation (Human Security Act of 2007). The ATA distinguishes between various types of terrorism, such as threats (any person threatening to commit a terrorist act), recruitment, incitement, and material support. Violations are tried in specific courts designated by the Supreme Court. The ATA also establishes an Anti-Terrorism Council, to be appointed directly by the President, overseeing terrorist investigations and with the power to "designate who is a "terrorist"" and authorise warrantless arrests.

Under the ATA, incitement of terrorism, "by mean of speeches, proclamations, writings, emblems, banners or other representations tending to the same end" is punished by life imprisonment without parole. The broad language used in the legislation allows for the penalisation of online material deemed to be inciting terrorism.

The legislation has been criticised by civil society and human rights groups – including Amnesty International and the Committee to Project Journalists – with critics warning that the ATA could be used to target government critics and that it threatens certain rights protected by the constitution. The main criticism concerns the criminalisation of incitement to terrorism, with critics concerned by the lack of clear definition of incitement which may provide "open-ended basis for prosecuting speech" as the Anti-Terrorism Council would be the unique arbiter of what constitutes a serious risk. The potential unconstitutionality of the law led to numerous petitions being signed to declare it unconstitutional. [15]

Opening an avenue for the criminalisation of terrorist use of the internet?

Even though the ATA does not explicitly address the issue of terrorist use of the internet, members of the Philippines' armed forces have raised the possibility of extending the bill to cover regulation of social media. Such actors argue that this expansion could help counter "preparatory acts" of terrorism, for example by facilitating the traceability of content that can be used for radicalisation or financing purposes. For now, the extension of the ATA to social media remains a suggestion. Former justice secretary Franklin Drillon has said that such an expansion would be in violation of the country's Bill of Rights, which enshrines freedom of information, whereas this law would impose regulation on service providers and social media platforms which might limit this.

Misuse of ICT

The legislation on Cybercrime Prevention acts as a framework for the use of information and communications technology (ICT) and lays out what are considered to be misuses and illegal access to ICT in the country, with individuals using online platforms being legally liable for content posted rather than tech companies. The legislation prohibits the use of ICT for any crimes penalised by the Philippines penal code and prohibits online defamatory content (known as "libel" content).

The prohibition of libel content online has led civil society groups to raise concerns about the law's potential impact on freedom of information and freedom of expression, in particular due to the lack of a precise definition of what constitutes online defamation. Furthermore, as highlighted by the Electronic Frontier Foundation, libel in the Philippines is not adjudicated on what the author meant, but rather adjudication is based on the meaning the words used *could* have. This leaves substantial room for interpretation, and therefore presents some risk for freedom of expression, as users might be tempted to self-censor to avoid potential judicial consequences.

A Magna Carta for Internet Freedom

One of the most noteworthy aspects of online regulation in the Philippines might not be the actual legislations, but rather the civic initiatives that surrounded them. Indeed, the drafting of the CPA led a group of citizens (or netizens, as they called themselves) to propose a Magna Carta for Philippine Internet Freedom (MCPIF). The Magna Carta was the product of a crowdsourcing effort from citizens concerned with digital rights. The MCPIF was drafted with participation from individuals with different

[15] 25 petitions as of 6 August 2020. At the time of writing, the signatories are waiting to be heard in an open argument by the Supreme Court in September 2020.

backgrounds – including IT specialists, bloggers, human rights advocates, lawyers, and academics – using online platforms to share their ideas and proposition for the enshrinement of digital rights in the Philippines.

Supported by Senator Miriam Defensor Santiago, the Magna Carta was created as a legislative response to the CPA that would ensure the protection of "the rights and freedoms of Filipinos in cyberspace", as well as provide definitions for certain cybercrimes. In particular, the MCPIF was drafted so as to protect netizens from illegal searches and prevent the blocking or restriction of a website with appropriate legal due process. Whilst the MCPIF did not succeed in replacing the CPA, the crowdsourcing effort behind it led to the creation of Democracy.net.ph, an "Internet and ICT Rights Advocacy Organization" that is still active today.

Resources

Republic of the Philippines (2020), *Anti-Terrorism Act.*

Republic of the Philippines (2012), *Cybercrime Prevention Act.*

Al-Jazeera (2020), *Philippine court asked to annul Duterte-backed anti-terror law.*

Amnesty International (2020), *Philippines: Dangerous anti-terror law yet another setback for human rights.*

Aspinwall (2020), *After Signing Anti-Terrorism Law, Duterte Names His Targets,* Foreign Policy.

Human Rights Watch (2020), *Philippines: New Anti-Terrorism Act Endangers Rights.*

Institute for Economics & Peace (2019), *Global Terrorism Index 2019: Measuring the Impact of Terrorism/*

Kemp (2020), *Digital 2020: The Philippines,* DataReportal.

Khaliq (2020), *Philippines: Anti-terror act faces top court challenges,* Anadalou Agency.

National Union of the Journalist of Philippines (2020), *No to criminalization of free speech.*

Pazzibugan (2020), *SC sets hearings on anti-terrorism law in September,* Inquirer.net

The Propinoy Project (2014), *Why there should be a Magna Carta For Philippine Internet Freedom.*

Ramos Christia M. (2020), "*Parlade defends social media regulation proposal", Inquirer*.net

Ramos (2020), *AFP Dials down Facebook 'regulation',* Inquirer.net.

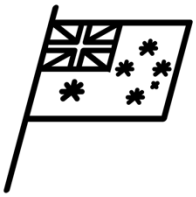
Reporters Without Borders (2012, updated in 2016), *Cybercrime law's threat to freedom of information.*

Sen. Santiago (2012), *Magna Carta For Internet Freedom To Replace Anti-Cybercrime Law,* Senate of the Philippines.

Yap, Protacio, Lopez, and Lazaro, (2020), *Anti-Terrorism Act signed into law,* Lexology.com

York (2012), *Philippines' New Cybercrime Prevention Act Troubling for Free Expression,* Electronic Frontier Foundation.

ASIA-PACIFIC | AUSTRALIA



Harmful and illegal online content have been regulated in Australia since the late-1990s via the Broadcasting Services Amendment (Online Services) Act of 1999, which established the legislative framework for online content regulation in the country.

Australia's regulatory framework:

- The *Online Content Scheme* (OCS), under Schedule 5 and 7 of the *Broadcasting Services Act July (*BSA), 1992, regulates "illegal and offensive" content in Australia.
- *Enhancing Online Safety Act 2015*, prohibits the sharing of, amongst other things, threatening posts on social media, and creates a "complaint and objection" system under the supervision of the newly established e-Safety Commissioner (2015).
- *The Telecommunications and Other Legislation Amendment (Assistance and Access) Bill,* 2018, enables law enforcement and intelligence agencies to require technical assistance from 'designated communications providers'.
- *The Criminal Code Amendment (Sharing of Abhorrent Violent Material),* Act 2019 creates two new types of offenses related to sharing of "abhorrent violent material" under the Criminal Code.
- *Online Safety Act proposal,* 2019, which sets out to reform and expand on existing online safety regulation.
- The *Online Safety Charter,* outlines Australia's expectations for online service providers to protect Australians from harmful online experiences.
- The Taskforce to Combat Terrorist and Extreme Violent Material Online, produced a report for government and industry on how to improve their ability to prevent and respond to future online crisis events. As a result of the report's recommendations, ISPs and the federal government have agreed to a new protocol to allow the blocking of websites hosting graphic material depicting a terrorist act or violent crime.
- Australia is a signatory of the Christchurch Call to Action.

Main bodies overseeing online regulation:

- The e-Safety Commissioner is empowered under the Enhancing Online Safety Act 2015.
    - The Commissioner administers the Online Content Scheme and can issue notices to service providers for content in violation of the Criminal Code Amendment Act 2019.
    - The Commissioner can also give written directions to ISPs to block Australian access to material that exposes the community to online terrorist and extreme violent material during crisis events.

Key takeaways for tech companies:

- All internet content and service providers operating in Australia are to comply with the *Online Content Scheme*, which provides a legal basis for prohibited online content.
- Violation of the *Criminal Code Amendment Act* 2019 – by either providing a content service or hosting service which can be used to access abhorrent violent material, and by failing to ensure expeditious removal or cease hosting of it following notification from authorities or failing to refer details to the Australian Federal Police after becoming aware of such content is available on their service – can be sanctioned by:
    - A fine of AU$2.1 million (around $1.5 million) or up to three years in prison (for an individual providing the content services or hosting services).
    - A fine up to AU$10.5 million or 10% of annual revenue for each offense (for a company).
- The e-Safety Commissioner can initiate investigations relating to online content and is able to take enforcement actions, such as by issuing notices:
    - The Commissioner can trigger the blocking of access in Australia to certain content hosted overseas by notifying the Australian ISPs of the content.
    - The e-Safety Commissioner can issue a notice, under the Criminal Code Amendment Act 2019, triggering the presumption that a service provider has been “reckless” about its service hosting abhorrent violent material.
- Tech companies should keep up to date with advancements on the Australian Government’s proposed Online Safety Act, which finished its consultation process in February 2020 and is pending a government response.

"Harmful and illegal" online content

Online content in Australia is regulated under Schedule 5 and 7 of Broadcasting Services Act, through the Online Content Scheme which establishes a complaints-based mechanism. Schedule 5 outlines provisions in relation to internet content hosted outside Australia, while Schedule 7 focusses on content services and user-generated content on the internet and mobile services hosted in or provided from Australia. Schedule 7 additionally defines "prohibited" or "potentially prohibited" content, where "prohibited" content is content that has been classified by the Classification Board as X18+ or RC (refused classification). Generally, the Online Content Scheme places restrictions on the types of online content that can be hosted or provided by internet service providers (ISPs) and content service providers.

Under the Enhancing Online Safety Act 2015, which promotes online safety for all Australians, the Online Content Scheme came into the administration of the newly established Australian e-Safety Commissioner. The e-Safety Commissioner oversees the regulation of access to illegal and "harmful" online content ("prohibited content" and "potential prohibited content").

The Commissioner responds to content complaints and has the capacity to initiate certain investigations relating to online content. The e-Safety Commissioner is able to take enforcement action, and indeed frequently reports particularly serious content to international law enforcement for investigation and removal. Although it cannot issue takedown notices for overseas hosted content, it can trigger the blocking of access to certain overseas hosted content through notifying Australian ISPs of the content. The Commissioner releases annual reports including evidence on their performance, key corporate information, and details against the mandatory reporting requirements. These reports include the number of investigations conducted into potentially prohibited online content, the number of URLs hosting material identified as likely to be prohibited, and the amount of notices to overseas services conducted in related to abhorrent violent material.

"Abhorrent violent material" and recklessness

Most recently, the Australian government enacted the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 – amending the 1995 Code. This amendment applies to ISPs, content service providers and hosting service providers anywhere in the world, including websites, social media platforms, and content management or cloud solution providers.

The Act creates two new types of offenses under the Criminal Code. The first is the failure by a service provider to notify the Australian Federal Policy within a "reasonable time" that "abhorrent violent material" relating to a conduct that is occurring in Australia is accessible on a service. According to the Act, the new category of content, "abhorrent violent conduct", refers to terrorism, murder, attempted murder, torture, rape or kidnapping. The second is the failure by a service provider to "expeditiously" remove, or cease to host, abhorrent violent material that is accessible within Australia. It is not clear how much time a platform has to comply with the legislation's requirements, since the Act does not define "expeditious". The Act further introduced the notion of "recklessness", stating that a company has been "reckless" if its service is being used to access or host abhorrent violent material.

The Criminal Code amendment empowers the e-Safety Commissioner to issue a notice triggering the presumption that a service provider has been "reckless" about its service hosting abhorrent violent material, unless the service provider can prove otherwise. According to Evelyn Douek, the law passed through both houses of parliament in a remarkably short time, limiting the possibility of any consultation from the industry or civil society. The UN special rapporteurs on counterterrorism and human rights and freedom of expression shared their comments on the law with the government via a letter, noting that the law raises serious concerns about freedom of expression, such as a consequence of imposing heavy fines and imprisonment on Internet intermediaries. The letter also addresses concerning ambiguities in the law, namely the definition of a "terrorist act" and "expeditiously".

Future proposals – The Online Safety Act

The Australian Government is currently looking into proposing a new Online Safety Act in order to reform and expand the online safety laws. This proposed Act would introduce a range of new aspects, including 24 hour take-down deadlines for harmful online content, and further empower the e-Safety Commissioner to direct Australian ISPs to block access to sites hosting terrorist or extreme violent material for a defined period where an online crisis event occurs.

The government also published an Online Safety Charter, defining their expectations of online service providers to protect Australians online, such as requesting service providers to take preventative steps to ensure that their service is less likely to facilitate, inflame or encourage illegal and inappropriate behaviours.

Legislation on Encryption

In 2018, the Australian federal parliament enacted the Telecommunications and Other Legislation Amendment (Assistance and Access) Act (TOLA), also known by the Media as the 'encryption laws'. This legislation enables law enforcement and intelligence agencies to require technical assistance from 'designated communications providers', encompassing 15 company types and spanning from social media companies to small hardware and software suppliers. The legislation permits a near unlimited range of technical assistance, going beyond decryption to include modifying consumer products and services.

However, the TOLA includes inspection and reporting requirements by the Commonwealth Ombudsman and the Home Affairs Minister. The Commonwealth Ombudsman may inspect the records of an interception agency to determine the extent of compliance by the agency and may conduct a written report to the Home Affairs on the results of the inspections. Furthermore, the Home Affairs Minister must prepare a written annual report that outlines the number of technical assistance requests and notices, as well as technical capability notices, that were given during the year by the chief of officers of interception agencies.

The legislation has been criticised by legal, civil society, and human rights organisations for having been enacted in a short timeframe and with little public consultation, as well as for its extensive powers, lack of clarity, and limited transparency requirements for the tech sector. Many groups

quickly wrote submissions to the Bill, including: Apple Inc., Australian Human Rights Commission, Australian Information Industry Association, Australian Information security Association, Digital Industry Group, and the Law Council of Australia. Despite the many issues raised, the Bill was passed in a single day. The enacted version of the Bill did include a list of amendments introduced by the government; however, none addressed the most commonly cited concerns, including the vagueness of possible technical assistance and a lack of judicial oversight.

Reviews into the TOLA have been conducted by the Parliamentary Joint Committee on Intelligence and Security (PJCIS) and the Independent National Security Legislation Monitor (INSLM). Although the INSLM has published its report, the PJCIS has yet to publish its review.

Resources

Australia Government (2019), *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act.*

Australia Government (2019), *Report of the Australian Taskforce to combat terrorist and extreme violent material online.*

Australia Government (2019), *Online Safety Charter.*

Australia Government (2019), *Consultation on Online Safety Reforms.*

Australia Government (2018), *Telecommunications and Other Legislation Amendment (Assistance and Access).*

Australia Government (2015), *Enhancing Online Safety Act.*

Australia Government (1992), *Broadcasting Services Act 1992.*

Australia eSafety Commissioner (2020), *ISP Blocking: Facts and Falsehoods.*

Arboleda, (2020), *Telcos, Govt Reach Agreement on How to Block Terrorist Content,* CRN.

Baker McKenzie (2019), *Unprecedented Penalties for Enabling the Sharing of Abhorrent Violent Material Online.*

Baker McKenzie (2020), *Australian Government Opens Consultation on New Online Safety Act/*

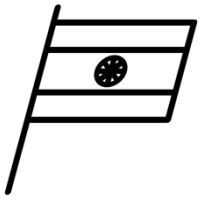
Australian Government (2019), *Online Safety Legislative Reform Discussion Paper*.

Douek (2019), *Australia's New Social Media Law Is a Mess ,* Lawfare Blog.

Kaye, Special Rapporteur on the Right to Freedom of Expression, Aolain, Special Rapporteur on Human Right and Fundamental Freedom while Countering Terrorism (2019), *Letter to the government of Australia.*

Hardy (2020), *Australia's encryption laws: practical need or political strategy?,* Internet Policy Review.

ASIA-PACIFIC | INDIA



With almost 500 million Internet users, and a history of mis- and disinformation spreading on social media and messaging apps and occasionally resulting in violence, content moderation has been a pressing issue in India for quite some time. Regulation of content is covered by different legislations under the Indian Penal Code, the Information Technology Act (ITA), and Criminal Procedure Code, and shortly under the *Framework and Guideline for use of Social Media*.

Terrorist use of the internet is mostly regulated through the criminalisation of cybercrime, covered by Section 66F of the Information Technology Act, which regulates cybercrimes and electronic commerce.

India's regulatory framework:

- The *Information Technology Act* (ITA), passed in June 2000 and amended in 2005, is the framework for regulating cybercrime, including the offence of cyberterrorism, in the country.
- *Shreya Singhal v. Union of India* (2012), a landmark decision by the Indian Supreme court in 2015 absolving tech companies from the obligation of actively monitoring their platforms for illegal content.
- *Framework and Guidelines for Social Media Regulations*, to be introduced in 2020, regulating the traceability of content shared on social media and messaging apps.

Key takeaways for tech companies:

- Tech platforms operating in India are exempted from liability for user-generated content, as long as they comply with government takedown guidelines regarding the removal of certain content, as per Section 79A of the ITA.
- Under the ITA, Section 69A, tech platforms can be asked to remove or block access to certain content deemed to be against the sovereignty, integrity and public order of India.
    - Non-compliance can be penalised by jail terms and fines.
- Under the *2020 Guidelines for of Social Media,* social media and messenger apps of over 5 million users will have to
    - Assist authorities tracking down the origin of a post within 72 hours of notice.
    - Keep data record of tracked content for 180 days.

ITA Section 66A vs. Indian Supreme Court

Whilst the ITA does not explicitly mention online terrorist content, Section 66A of the ITA penalised – until 2015 – the use of communication devices to share information that is "grossly offensive or has menacing character", as well as false information purposely shared to cause (amongst other

things) hatred, ill will, criminal intimidation, and enmity. The provision was struck down by the Supreme Court of India in *Shreya Singhal v. Union of India* (2012). The Court held that the prohibition against the dissemination of false information did not fall within any reasonable exceptions to the exercise of the right to freedom of expression. Instead, it held that online providers would only be obligated to take down content upon receiving an order from a court or government authority. The court further clarified that the "public order" free speech exception under Article 19(2) of the Constitution would not apply to cases of "advocacy" but only to "incitement", which has a proximate relation to public disorder.

In effect, this decision by the Court exempted tech platforms from having to pro-actively monitor content and exempted them from liability for user-generated content. However, the Court also "made it clear that only authorised government agencies and the judiciary could legitimately request internet platforms to take down content" that is deemed illegal by the Indian authorities.

A content regulation system based on removal and blocking requests

Without a regulatory framework dedicated to online content and speech, the Indian government has relied on content takedown and blocking requests to moderate online content. Under Section 69A of the ITA, Indian authorities can request that tech platforms block access to content in the "interest of sovereignty and integrity of India or public order or for preventing incitement to the commission of any cognizable offence relating to the above". Moreover, tech platforms' exemption from liability for user-generated content is linked to removal or blocking of access to content when notified by the Indian authorities, as per Section 79A. Those failing to comply with a blocking request can face up to seven years of imprisonment and a fine.

This reliance on removal requests has made India the leading country in government content removal requests sent to major tech platforms, in particular Facebook, according to a 2019 report by Comparitech. Most of the requests made by the Indian authorities were related to "hate speech, anti-religion content constituting incitement to violence, extremism, and anti-state content." The Indian government has also been relying on internet shutdown measures, with 134 reported shutdowns in 2018. These shutdowns range from long-term shutdowns in places like Kashmir to more sporadic, short-term shutdowns in response to protests and unrest in the country.

"Traceability" of content

In response to the misuse of social media platforms and messaging apps, and in particular to counter the spread of mis- and disinformation, the Ministry of Electronics and Information Technology introduced in February 2020 a *Framework and Guidelines for of Social Media Regulations.* This framework was first introduced by the Ministry of Electronics and Information Technology in January 2018, and was set to be rolled out in 2020. This framework would allow for facilitated sharing of information between tech companies and law enforcement agencies, and require platforms to help authorities track down the origin of any post within 72 hours. Companies would also have to "keep records [of tracked online content] on file for 180 days at minimum to aid with potential government investigations", as well as to maintain a physical presence in the country and establish a "grievance officer" to liaise with the government.

These rules would apply to all social media and messaging apps with more than 5 million users in India, whilst other tech companies, such as operating systems or online encyclopaedias and repositories, are all exempt. Encrypted messaging apps would also have to comply with the treatability requirement, a task complicated to complete without breaking the end-to-end encryption. Tech platforms and civil rights advocates have criticized the new rules for being "an invitation to abuse and censorship", as well as a burdensome requirement for companies to comply with. Others, such as the Internet and Mobile Association of India – including Facebook and Alphabet – have criticized the rules for being detrimental to the right of privacy.

Resources

Awashti (2020), *Social media users to be tracked by government under new guidelines: Report,* The Hindu Business Line.

BBC News (2018), *India lynchings: WhatsApp sets new rules after mob killings.*

Bischoff Paul (2019), *Which government censors the tech giants the most?,* Comparitech.

Bristows LLP (2020), *Social media; to regulate or not to regulate?,* Lexology.

Cloudflare, *GitHub and Mozilla's open letter to the Indian Government.*

Library of Congress, *Government Responses to Disinformation on Social Media Platforms: India.*

Mandavia (2019), *India sent most takedown requests to social media companies*, Economic Times India.

Nazmi (2019), *Why India shuts down the internet more than any other democracy*, BBC News.

Newton (2020), *India's proposed internet regulations could threaten privacy everywhere*, The Verge.

PYMNTS.com (2020), *India's New Social Media Rules Would Strip Anonymity — When Asked — From Accounts.*

Rai (2020), *400 million social media users are set to lose their anonymity in India,* Bloomberg.

Taneja and Shah (2019), *Kashmir blackout: Counterterrorism and an increasingly challenging role of the internet,* Observer Research Foundation.

Samuel (2020), *How misinformation on WhatsApp led to a mob killing in India*, The Washington Post.

Sidharthan R., *The Information Technology Act and Media Law,* Legal Service India.

Sidharthan and Awasthi (2020), *Social media users to be tracked by government under new guidelines,* The Hindu Business Line.

Software Freedom Law Center (2019), Any regulation of online speech in India must safeguard the rights to free speech and privacy, Scroll.in.

United Nations Office on Drugs and Crime (2012), *Use of the Internet for Terrorist Purposes.*

Wagner (2019), *WhatsApp is at risk in India. So are free speech and encryption.*, Vox.

NORTH AMERICA | THE UNITED STATES

Online regulation and content moderation in the United States is defined by the First Amendment right to freedom of speech and Section 230 of the Communication Decency Act 1996, which establishes a unique level of immunity from legal liability for tech platforms. It has broadly impacted the innovation of the modern Internet, causing global effects beyond the US. Recently, however, the Trump Administration administered an executive order directing independent rules-making agencies to consider regulations that narrow the scope of Section 230 and investigate companies engaging in "unfair or deceptive" content moderation practices. This shook the online regulation framework and resulted in a wave of proposed bills and Section 230 amendments from both government and civil society.

US' regulatory framework:

- *First Amendment law under the US Constitution* outlines the right to freedom of speech for individuals and prevents the government from infringing on this right, for example by banning certain types of speech.
- *Section 230 of the Communication Decency Act of 1996* establishes intermediary liability protections related to user-generated content in the US, meaning that tech companies are not seen as liable for content posted by their users.

Relevant national bodies:

- The Federal Communications Commission (FCC) regulates interstate and international communications by radio, television, wire, satellite and cable in all 50 states, the District of Columbia and US territories.
    - An independent US government agency overseen by Congress, the commission is the primary domestic authority for communications law, regulation and technological innovation.

Key takeaways for tech companies:

- First Amendment law establishes Internet platforms as being in control of their own content policies and codes of conduct.
- Under Section 230, web hosts, social media networks, website operators, and other intermediaries are largely shielded from being held liable for user-generated content. Companies are able to moderate content on their platforms without being held accountable.
- However, this might change soon. There are currently two bipartisan bills for Section 230 which experts say have a chance of passing:
    - The Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2019 (EARN IT Act), introduced in March 2020. Under this Act, companies would have to "earn" Section 230 immunity based on their content moderation practices.
    - The Platform Accountability and Consumer Transparency Act (PACT Act), introduced in June 2020, focuses on promoting platform transparency and accountability.
- Further, President Trump issued an Executive Order in May 2020 in which he directed independent rules-making agencies, including the FCC, to consider regulations that narrow the scope of Section 230 and investigate companies engaging in "unfair or deceptive" content moderation practices.

Freedom of expression online

Legally speaking, regulation of online content and of content moderation practices by technology companies operating in the US has been limited to date. This is due to two principal legal frameworks that shape the US' freedom of expression online: The First Amendment to the US Constitution and Section 230 of the Communications Decency Act (CDA).

The First Amendment outlines the right to freedom of speech for individuals and prevents the government from infringing on this right. Internet platforms are able to establish their own content policies and codes of conduct. Section 230 of the Communication Decency Act of 1996 (CDA) establishes intermediary liability protections related to user-generated content in the US. The broad immunity granted to technology companies in Section 230 states that "no provider or user of an interactive computer service shall be treated as a publisher or speaker of any information provided by another information content provider." Companies are, therefore, able to moderate content on their platforms without being held accountable. In other words, online platforms have the freedom to police their sites and restrict material as they see fit, even if speech is constitutionally protected. For example, this protects platforms from lawsuits if a user posts something illegal, although there are exceptions for copyright violations, sex-work related material, and violations of federal criminal law. It is important to note that Section 230 of the CDA is unique to American law: European countries, Canada, Japan, and the vast majority of other countries do not have similar statutes on their books.

The historical context behind Section 230 is complex, but it gives an illuminating look into the culture of free speech in the US and its relation to content online. The statute was the product of debates over pornography and other "obscene" materials in the early 1990s. With the advent of early internet

services like CompuServe or Prodigy, US Courts tried to understand whether those service providers were to be treated as "bookstores" (neutral distributors of information) or as "publishers" (editors of that information) when adjudicating their standing under the First Amendment. A court ruled that CompuServe was immune to liability because it was similar to a bookstore, while Prodigy did not get the same immunity due to its enforcement of its own content moderation policies – thereby, making it a publisher. In other words, companies were incentivised to not engage in content moderation in order to preserve their immunity. Section 230 of the CDA sought to change this mismatch of incentives by preserving the immunity of these platforms and providers while they engage in content moderation.

Recent Amendments

The question of content moderation has to some extent developed into a partisan cleavage between the liberal Democratic Party and the conservative Republican Party in recent years. Democrats tend to claim that online platforms do not moderate enough and are therefore complicit in the spread of hate speech and disinformation. Republicans, on the other hand, often argue that these companies moderate too much, producing an alleged 'liberal bias' that they say undermines 'conservative' content. As a result, there has been a flurry of recent legislative and executive proposals to influence content moderation.

In June 2019, Republican? Senator Josh Hawley introduced the "Ending Support for Internet Censorship Act," which seeks to amend Section 230 so that larger internet platforms may only receive liability protections if they are able to demonstrate to the Federal Trade Commission that they are "politically neutral" platforms. However, the Act raises First Amendment concerns, as it tasks the government to regulate what platforms can and cannot remove from their websites and requires platforms to meet a broad, undefined definition of "politically neutral."

President Trump issued an executive order in May 2020 directing independent rules-making agencies, including the Federal Communications Commission, to consider regulations that narrow the scope of Section 230 and investigate companies engaging in "unfair or deceptive" content moderation practices.[16]

On June 17 this year, Senator Josh Hawley (R-MO) most recently introduced the Section 230 Immunity to Good Samaritans Act. Nominally, the Hawley bill would prevent major online companies from receiving the protections of Section 230 of the CDA unless their terms of service were revised to operate "in good faith" and publicise content moderation policies. According to Senator Hawley, "the duty of good faith would contractually prohibit Big Tech from discriminating when enforcing the terms of service they write and failing to honor their promises". This would open companies to being sued for breaching their contractual duties, along with a fine of $5,000 per claim or actual damages, whichever is higher, in addition to attorney's fees.

[16] Critics have underlined that the enforcement of this order is legally debatable and raises questions regarding the administration's approach to regulating content moderation, given the First Amendment protections do not allow anyone to determine what a private company can or cannot express. See: Why Trump's online platform executive order is misguided, Brookings, Niam Yaraghi.

Following President Trump's executive order, the Department of Justice issued a proposal in September for legislatively rolling back Section 230. This draft legislation focuses on two areas of reform, which, according to the DOJ are "necessary to recalibrate the outdated immunity of Section": promoting transparency and open discourse; and addressing illicit activity online. The DOJ also shared their own recommendations for altering Section 230 with Congress. If enacted, the DOJ recommendations would pave the way for the government to impose steep sanctions on platforms if they do not move to remove illicit content, including that related to terrorism.

Bipartisan bills

According to an evaluation of the proposed Section 230 bills by Paul M. Barrett, the deputy director of the NYU Stern Center for Business and Human Rights, two bipartisan Senate bills "have at least a chance of eventual passage": the EARN IT Act and the PACT Act.

- The Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2019, proposed by Senators Lindsey Graham (R-SC) and Richard Blumenthal (D-CT) in March 2020: The general idea behind the EARNT IT Act is that tech companies will have to "earn" Section 230 immunity based on their content moderation practices, rather than being granted immunity by default. The bill was proposed by lawmakers as a way to counter child sexual abuse material (CSAM). To earn Section 230 protections, the bill in March introduced a National Commission on Online Child Sexual Exploitation Prevention that would set content moderation standards for tech companies to meet. Amendments were made to the bill in July this year, including that the standards set by the commission would not be requirements, but instead voluntary recommendations. However, the changed bill would still allow states to sue tech platforms if child sexual abuse material appears on their platforms. Critics say that the bill poses a threat to Section 230 protections and encryption. For example, if child abuse material is sent through an encrypted messaging platform, states will be able to sue the platform and hold them responsible for being unable to moderate those messages. The Senate Judiciary Committee voted to approve the EARN IT Act for a floor vote on July 2, 2020. According to the Electronic Frontier Foundation (EFF), the EARN IT Act passed the Senate Judiciary Committee in September, and has since been introduced in the House of Representatives.

- The Platform Accountability and Consumer Transparency (PACT) Act, introduced by Senators Whip John Thune (R., S.D.) and Senator Brian Schatz (D., Hawaii) in June 2020. The PACT Act focuses on promoting platform transparency and accountability. The Act includes a requirement that platforms explain their content moderation policies to users and provide detailed quarterly statistics on items removed, down-ranked, or demonetised. It would amend Section 230 to give larger platforms just 24 hours to remove content that is determined unlawful by a court. The platforms would also have to develop a complaint system that notifies users within 14 days of takedowns and to provide for appeals. Another part of the Act would allow federal regulators to bring civil enforcement lawsuits against platforms. According to Access Now's assessment of the PACT Act, the Act's notice-and-takedown mechanism "lacks critical safeguards and clearer procedural provisions", but this proposal "has the potential to serve as a valuable framework with some restructuring and tweaks".

Beyond Government – Scholars and Civil Society

Scholars and civil society have developed their own reports and recommendations to amend Section 230, and some have even proposed entirely new regulatory frameworks and agencies to oversee US content moderation.

Beside government proposals, a 2019 report, published by the University of Chicago's Booth School of Business, suggests transforming Section 230 into a "quid pro quo benefit." Platforms would have a choice: adopt additional duties related to content moderation or forgo some or all of the protections afforded by Section 230.

Another proposal comes from Danielle K. Citron, a law professor at Boston University. Citron has suggested to amend Section 230 by including a "reasonableness" standard, which would mean conditioning immunity on "reasonable content moderation practices rather than the free pass that exists today". The "reasonableness" would be determined by a judge at a preliminary stage of a lawsuit, wherein the judge would assess the "reasonableness" of a platform's overall policies and practices.

Regulatory framework proposals beyond Section 230

Others have yet studied another idea: the creation of a new federal agency specifically designed to oversee digital platforms. A study released in August 2020 by the Harvard Kennedy School's Shorenstein Center on Media, Politics, and Public Policy proposes the formation of a Digital Platform Agency. The study recommends that the agency focus on promoting competition among internet companies and protecting consumers in connection with issues such as data privacy.

In a report, The Transatlantic Working Group (TWG) has emphasised the need for a flexible oversight model, in which authorising legislation could extend the jurisdiction of existing agencies or create new ones. As possible examples of existing agencies, the TWG cites the US Federal Trade Commission, the French Conseil Supérieur de L'Audiovisuel, and the British Office of Communications, or OFCOM. The TWG overlaps in some of the goals of the PACT Act, for instance in requesting greater transparency. The TWG envisions a digital regulatory body that requires internet companies to disclose their terms of service and their enforcement mechanisms.

Resources

Barrett (2020a), *Regulating Social Media: The Fight Over Section 230 — and Beyond*, NYU Stern.

Barrett (2020b), *Why the Most Controversial US Internet Law is Worth Saving*, MIT Technology Review.

Brody and Null (2020), *Unpacking the PACT Act,* Access Now.

Feiner (2020), *GOP Sen. Hawley unveils his latest attack on tech's liability shield in new bill,* CNBC.

Hawley (2020), *Senator Hawley Announces Bill Empowering Americans to Sue Big Tech Companies Acting in Bad Faith.*

Mullin (2020), *Urgent: EARN IT Act Introduced in House of Representatives*, Electronic Frontier Foundation.

Newton (2020), *Everything You Need to Know About Section 230*, The Verge.

New America (2019), *Bill Purporting to End Internet Censorship Would Actually Threaten Free Expression Online.*

Ng (2020), *Why Your Privacy Could be Threatened by a Bill to Protect Children,* CNET.

Robertson (2019), *Why the Internet's Most Important Law Exists and How People Are Still Getting it Wrong*, The Verge.

Singh (2019), *Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, New America.

Yaraghi (2020), *Why Trump's online platform executive order is misguided*, Brookings.

Government of the United States, *The Constitution,* Whitehouse.gov

Government of the United States (2020), *Executive Order on Preventing Online Censorhip,* Whitehouse.gov

Legal Information Institute – Cornell Law School, *47 U.S. Code § 230 - Protection for private blocking and screening of offensive material.*

Congress of the United States*, The Eliminating Abusive and Rampant Neglect of Interactive Technologies Act 2019.*

Congress of the United States *, Platform Accountability and Consumer Transparency 2020*

NORTH AMERICA | CANADA

Canada's approach to online regulation has, so far, been characterised by its support for tech sector self-regulation as opposed to government-led regulation of online content. However, concerns over foreign interference in Canadian politics and online hate speech and extremism, have led to public discussions considering the introduction of a legislation on harmful online content, and the possibility to make tech companies liable for content shared on their platforms.

Canada's regulatory framework:

- *National Strategy on Countering Radicalization to Violence*, 2018, which summarises Canada's approach to countering terrorism and violent extremism.
- *Canada's Communications Future: Time to Act,* (BTLR), January 2020, a broad review of the broadcasting and telecommunications legislation in Canada, drawing recommendations for the future of the legislative framework in the country, and calling for the introduction of social media regulation.
- *Canada's Digital Charter,* 2019, which lays out Canada's approach to internet technologies and the online space; with the 9th principle addressing the issue of violent extremism, and underlining that the online space should be "free from hate and violent extremism".
- *Digital Citizen Initiative,* Canada's strategy for the building "resilience against online disinformation and […] support a healthy information system", focused on research and "citizen" activities.
- Canada is a signatory to the Christchurch Call to Action.

Main regulatory bodies:

- Canadian Radio-television and Telecommunications Commission, which oversees the regulation of internet services in the country.
- Public Safety Canada (Ministry of Public Safety and Emergency Preparedness) – the main federal body in charge of coordinating matters related to national security, safety and maintaining a peaceful society.
    - Canada Centre for Community Engagement and Prevention of Violence, responsible for the *National Strategy on Countering Radicalization to Violence.*
- Innovation, Science and Economic Development Canada, which oversees different areas of Canada's economic development, published the 2020 broadcasting and telecommunications legislative review.
- Canadian Heritage, which oversees the *Digital Citizen Initiative*.
    - In December 2019, Steven Guilbeault, the newly appointed Minister of Canadian Heritage was tasked, by Prime Minister Justin Trudeau, in his Mandate Letter to develop a new regulatory framework for social media: "starting with a requirement that all platforms remove illegal content, including hate speech, within 24 hours or face significant penalties. This should include other online harms such as radicalization, incitement to violence, exploitation of children, or creation or distribution of terrorist propaganda."

Key takeaways for tech platforms:

- Tech platforms are exempt from liability for user-generated content.
- Canada has favoured a self-regulatory approach to moderation of online content and speech, engaging in cross-sector initiatives to support the tech sector in countering terrorist and violence extremist use of the internet.
- The *Canada's Communications Future: Time to Act* (2020), known as BTLR, offers a blueprint for regulating online content in the country, calling for tech companies to be held liable for harmful content on their platforms.[17]

Support for self-regulation and cross-sector initiatives

Both the Digital Charter and National Strategy to counter radicalisation stresses that citizens should be able to "fully participate in the online spaces" without viewing harmful and extremist content. Further, Canada's framework to counter terrorism and extremism has comprehensively integrated the need to tackle terrorist use of the internet. The 2018 National Strategy lays out the principle of Canada's approach to extremist content online, which is based on a "multi-stakeholder approach that includes national and international engagement with technology companies, academic researchers and civil society."

[17] At the time of writing, there are still uncertainties about whether the recommendations made in the BTLR are to become laws in Canada.

This has led Canada to focus on digital literacy and counter narratives efforts and on supporting research efforts to better comprehend the terrorist and violent extremist online landscape in the country. Most of these initiatives are funded via the *Community Resilience Fund*.

Supporting innovative tools for a swifter identification of terrorist content

Following Canada's signing of the Christchurch Call to Action, Public Safety Canada announced that it would award a grant to Tech Against Terrorism to develop the Terrorist Content Analytics Platform (TCAP). The TCAP will be the world's largest database of verified terrorist content, aimed at supporting tech companies in swiftly identifying terrorist content uploaded on their platforms, and will inform quantitative research on terrorist use of the internet.

Canada's support for the TCAP demonstrates the country's acknowledgment of the difficulties faced by small and micro tech companies in tackling terrorist exploitation, and of its willingness to support content moderation via innovative tools.

BTLR: towards regulation of online content and speech?

Concerns about online foreign interference in Canadian's politics and elections has led to calls for regulating online content, especially on social media. Former Minister of Democratic Institutions, Karina Gould, called for such regulation, arguing that tech platforms were demonstrating a "lack of willingness" from tech companies to address the issue.

In addition, concerns for the future of the Canadian digital space, including extremist and harmful content, has led Canada to consider regulatory approaches. In June 2018, the government commissioned a legal review of the communication legislative framework, which resulted in the *Canada's Communications Future: Time to Act* report. The report highlights concerns related to the spread of harmful content and extremist views online. Mainly, it recommends the introduction of a "legislation with respect to liability of digital providers for harmful content and conduct using digital technologies." Such legislation would aim to counter the spread and amplification of "harmful content" (a term which remains undefined in the report) online.

With regards to illegal content – including terrorist content – the BTLR recommends that the Canadian government introduces regular reviews of tech platforms' monitoring and removing mechanisms for "illegal content and conduct found online".

Further, the BTLR also recommends the establishment of a registration system for tech companies operating in Canada, which would bring all media content providers under a newly formed "Canadian Communications Commission". Registered companies would then have to "provide such information as the CRTC [Communications Commissions] may specify, " and will be obliged to support the diffusion of Canadian content. This registration would differentiate between different type of media providers: content curation, such as Netflix or Spotify; content sharing, such as Facebook and YouTube, and content aggregation for media disseminating content from curators, which would mostly apply to traditional media broadcast services.

Whilst many of the regulatory proposals to regulate tech platforms and online content are still in an early phase, online regulation in Canada is likely to undergo major changes and to see the embedding of the principle of legal liability for user-generated content for tech platforms in the Canadian online landscape.

Resources

Austen (2019), *Canada Joins the World in a Social Media Crackdown,* The New York Times.

Baker McKenzie (2018), *Government of Canada Looks to Modernize Telecommunications and Broadcasting Legislation for the Digital Age*, Lexology.

Boutilier, Oved, Silverman, and L. Jane (2019, updated in 2020), *Canadian government says it's considering regulating Facebook and other social media giants,* The Hamilton Spectator.

Jeftovic (2020), *Canada's BTLR is a blueprint for regulating internet* content, Easydnscom.

Canada Centre for Community Engagement and Prevention of Violence (2018), *National Strategy on Countering Radicalization to Violence.*

Government of Canada (2019), *Canada Declaration on Electoral Integrity Online.*

The Guardian (2019), *Canada may regulate social media companies to avoid election meddling.*

Innovation, Science and Economic Development Canada (2019), *Canada's Digital Charter in Action: A Plan by Canadians, for Canadians.*

Innovation, Science and Economic Development Canada (2019), *Canada's Communications Future: time to act" Broadcasting and Telecommunications Legislative Review.*

Library of Congress, *Government Responses to Disinformation on Social Media Platforms: Canada.*

OpenMedia (2020), *The BTL…What? What is the BTLR report and what it means for the future of our Internet.*

Government of Canada – Innovation, Science and Economic Development Canada (2019), *Canada's Digital Charter in Action: A Plan by Canadians, for Canadians.*

Government of Canada – Innovation, Science and Economic Development Canada (2020), *Canada's Communication Future: Time to Act.*

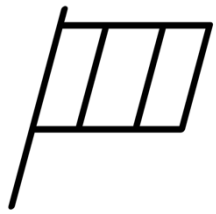
Public Safety Canada (2019b), *Government of Canada Announces Initiatives to Address Violent Extremist and Terrorist Content Online.*

Government of Canada – Office of the Primer Minister (2019), *Minister of Canadian Heritage Mandate Letter.*

Government of Canada – Canadian Heritage, *Online Disinformation.*

Government of Canada – Canada Centre for Community Engagement and Prevention of Violence (2018), *National Strategy on Countering Radicalization to Violence.*

EUROPE | FRANCE



France is, alongside New Zealand, an initiator of the *Christchurch Call to Action* to eliminate terrorist and violent extremist content online. Prior to the Christchurch Call, France has elevated tackling terrorist use of the internet as a key pillar of its counterterrorism policy,[18] supporting the EU proposal on Preventing the Dissemination of Terrorist Content Online, including the requirement for tech platforms to remove flagged terrorist content within one hour.

France's regulatory framework:

- *Countering online hate law*. Adopted in May-June 2020, the so-called "cyber-hate" or "Avia" law[19] establishes France's new broad framework to counter hateful, discriminatory, terrorist, and child sexual abuse (CSA) content online – all of which are illegal under French law.
    - The law would compel companies to remove terrorist and CSA content within one hour of being notified by French authorities, and within 24 hours for hateful and discriminatory content.
    - Following a "censuring" by the French Constitutional Council, which deemed the law to be bearing disproportionate risks to freedom of expression, the removal requirement was lifted and the law is now reduced to its preventive component.
- *Law on strengthening the provisions relating to the fight against terrorism,* November 2014, strengthens France's counterterrorism approach and introduces the penalisation of "terrorism apology" (apologie du terrorisme) and incitement, including for content shared online.
- France is a signatory and co-initiator of the *Christchurch Call to Action*.

[18] "La lutte contre l'utilisation d'internet à des fins terroristes constitue l'un des axes majeurs de l'action de la France en matière de contre-terrorisme."
[19] In France it is common practice to nickname a law with the last name of the political figure who proposed it to parliament, in that case MP Laetitia Avia from *La République en Marche*.

Main regulatory bodies:

- Conseil Superieur de l'Audiovisuel (CSA), an independent body which oversees broadcast communications (TV and radio) in France:
    - Under the new "cyber-hate" law, the CSA will coordinate an "Online Hate Observatory" to analyse the spread of hate online.
- Ministry of Interior, oversees – alongside judicial authorities – reports of terrorism apology and incitement, including for online content.
    - Manages *Pharos,* France's online content reporting platform.
    - Office central de lutte contre la criminalité liée aux technologies de l'information et de la communication (Cybercrime unit) which takes part in the coordination of content reported via Pharos and liaise with Europol's Internet Referral Unit.
- State Secretary for Digital Affairs, coordinates France digital policy and related discussion on the on the online regulatory framework (both at the national and international level).
- Digital Ambassador, coordinates international digital policy and transformation issues, including cyber security and online regulation.

Key takeaways for tech platforms:

- Despite recent attempts, including the "cyber-hate" law, France does not currently regulate tech platforms.
- However, certain content is considered illegal under French law, including terrorist (incitement and apology) content.
    - French authorities can require a website to be blocked or a piece of content to be removed if terrorist content is located
    - Authorities can require that a website or piece of content is removed from French search engine results
    - Individuals posting terrorist content risks seven years' imprisonment and a 100,000 euro fine[20]
- Internet users can report illegal content to the French authorities via Pharos, a platform dedicated to user reporting of illegal content online.

[20] Marine Le Pen, the leader of far-right *Rassemblement National*, a French MP, and former Presidential candidate and EU MP, has been tried for sharing Islamic State execution photos on Twitter in 2015.

Towards a more stringent framework?

To complement France's 2014 legal framework on online terrorist content, a law on countering online hate was submitted to Parliament on 20 March 2019, only a few days after the Christchurch shooting. The "cyber-hate" law (also known as the Avia law) was passed on 13 May 2020.[21]
Similar to the EU proposition for preventing the dissemination of terrorist content online – which would require tech platforms to remove content terrorist content within one hour, our blog post here – the Avia law had a its core a requirement for tech platforms to remove illegal content or face a substantial fine of up to 4% of the platform's annual global turnover. Under this law, terrorist and child sexual abuse material would have had to be removed within one hour of notification by the French authorities; and any other harmful content, as defined by existing French Law, (including incitation to hatred) would have had to be removed within 24 hours of flagging by any user. However, the removal deadline were censured by the French Constitutional Council which stripped the law of all its requirements for tech companies[22], keeping only its preventive aspect and calling for increased transparency and accountability from the tech sector, though without specifying what this entails exactly. The final version of the law also maintains the establishment of an "Online Hate Observatory" overseeing the enforcement of the law and will publish an annual report on it.

Censuring by the French Constitutional Council: risks for freedom of expression

In censuring the law, the Constitutional Council raised a number of concerns related to the risks of over-censoring online content, and on the potential impossibility for platforms, particularly smaller companies, "to satisfy" the removal requirements. Its ruling also underlines that the decision to adjudicate on illegal content and thus of what constitutes a valid limit to freedom of expression – which platforms would have done by removing illegal content within a short-removal period without judicial oversight – should not belong to tech platforms but remain a judicial decision inscribed in the rule of law.

Below, we summarise some of the most important concerns raised with regards to the Avia Law, and the Constitutional Council's arguments to censur them.

- A lack of consideration for smaller tech platform capacity. A one-hour delay for removal of terrorist content is unrealistic for micro and small platforms which lack the necessary human and technical resources to respond within such a short deadline. A one-hour time period, and even a 24-hour one for other hateful and discriminatory content, would require constant monitoring from tech platforms to ensure compliance, which would prove difficult, if not impossible, for most tech platforms. The French Constitutional Council particularly underlined that the law included regulations that were "impossible to satisfy" for tech platforms, thus breaking the principle of equality with regard to public regulations.

[21] On the law legislative timeline, it should be noted that the proposal benefited from an accelerated process granted by the government on 2 May 2019. When passed, it became the first non Covid-19 related law to be passed in the country since early March 2020, only two days after the lockdown was lifted in the country. This has led to some commentaries regarding the French government using the wave of misinformation linked to the pandemic as "the perfect impetus" to have it passed despite its critics.
[22] Beside the removal requirements, the law also required tech companies to blocking of mirror sites, and de-referencing of such sites on search platforms, as well as to designate a physical person located on French territory to act as a focal point and receive removal notifications.

- Risks for freedom of expression. Due to the short deadline and the broad scope of the law, tech platforms would not have had the time to properly adjudicate on a piece of content's legality. This could promote overzealous removal of content, with platforms indiscriminately taking down all content notified (without assessing whether it is, in fact, illegal) and increasingly relying on automated moderation tools, to ensure that they do not get fined before . Whilst automated moderation has its benefits, many solutions lack nuance and require human overview to avoid the excessive takedown of content. An overreliance on such methods presents risks for freedom of expression as it could lead to taking down lawful content.[23] This was stressed by the French Constitutional Council, which deemed that the removal requirement were neither necessary, appropriate or proportionate.

- Leaving tech platforms to adjudicate on illegality. The law itself did not create a new set of harms, nor did it create a new range of prohibited content. Everything, from hateful and discriminatory to terrorist and child sexual abuse content, is already illegal under French law. However, the legal definitions of such content are broad, and limitations to freedom of expression[24] have to be decided by an independent judiciary body, such as a judicial court. This is problematic since the law places responsibility to (rapidly) decide what is hateful or discriminatory content to private tech companies. In effect, this could lead to a development where private tech companies decide on what content is illegal according to their interpretation of the law instead of adequate legal channels. In this regard, the Constitutional Council's decision was a strong reminder that adjudicating on the legality of an online content, in particular terrorist content, is "subject to the sole discretion of the [French] administration."

Resources

Berne (2016), *Dans les coulisses de la plateforme de signalement Pharos,* NextInpact.

Breeden 2020), *French court strikes down most of online hate speech law*, The New York Times.

Chandler (2020), *France social media law is another coronavirus blow to freedom of speech,* Forbes.

Hadavas (2020), *France's New Online Hate Speech Law Is Fundamentally Flawed*, Slate/

La Maison des Journalistes, *Les limites de la liberté d'Expression.*

L'Express (2019), *Marine Le Pen renvoyée en correctionnelle pour avoir posté des images d'exactions de l'EI.*

Lapowsky (2020), *After sending content moderators home, YouTube doubled its video removals, Protocol.*

Lausson (2020a), *Très contestée, la « loi Avia » contre la cyberhaine devient réalité,* Numerama.

[23] On that, it is interesting to note that increased reliance on automated moderation has led to different results for Facebook and Youtube, both had to reduce human moderation during the lockdowns ensuing the Covid-19 crisis. Whilst this led to less content being taken down on Facebook, moderators were not able to log content into the automated system, Youtube had doubles its removals as it increased its reliance on automation.

[24] In France, freedom of expression, whilst protected by Article 11 of the *Declaration of the Rights of Man and of the Citizen of 1789,* is not absolute and can be limited. French law notably prohibits incitement to racial, ethnic or religious hatred, glorification of war crimes, discriminatory language on the grounds of sexual orientation or disability, incitement to the use of narcotics, Holocaust denial.

Lausson (2020b), *La loi Avia contre la haine sur Internet s'effondre quasi intégralement,* Numerama.

Ministère de l'Europe et des Affaires Etrangères. – France Diplomatie (2019), *Réguler les contenus diffusés sur l'internet et régulation des plateformes*".

Ministère de l'Europe et des Affaires Etrangères. – France Diplomatie, (2020), *Gouvernance d'Internet, quels enjeux ?*

Pielemeier and Sheehy (2019), *Understanding the Human Rights Risks Associated with Internet Referral Units,* The Global Network Initiative Blog.

Schulz (2020), *What's Going on With France's Online Hate Speech Law*?, Lawfare.

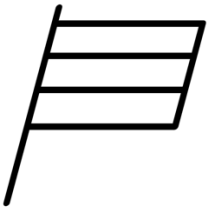
Kaye (2019), *Mandat du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression.*

République Française – Service-publique.fr (2020), *Incitation à la haine, à la violence ou à la discrimination raciale.*

République Française – Légifrance, *Loi n° 2014-1353 du 13 novembre 2014 renforçant les dispositions relatives à la lutte contre le terrorisme.*

Conseil Constitutionnel (2020), *Décision n° 2020-801 DC du 18 juin 2020 : Loi visant à lutter contre les contenus haineux sur internet.*

EUROPE | GERMANY



Germany has an extensive framework for regulating online content, particularly with regards to hate speech and violent extremist and terrorist material. Experts also note that Germany's regulatory framework has to some extent helped set the standard for the European, and possibly global, regulatory landscape.

Germany's Regulatory Framework:

- *The Network Enforcement Act (NetzDG),* June 2017, aims to counter 22 different online offences, including cyberbullying, disinformation, child sexual exploitation, defamation, and terrorist use of the internet.
- *The Gesetzentwurf zur Bekämpfung des Rechtsextremismus und der Hasskriminalität,* also called the February 2020 amendment to the NetzDG, aims to counter right-wing extremism. This amendment is currently on hold as there are concerns that it might be unconstitutional.
- The April 2020 amendment, furthers the requirements put on tech companies in the NetzDG and adopts the obligations set in the European Union's Audio-Visual Media Services Directive (AVMSD) 2018 into national law[25]. Germany's obligations under the AVMSD will therefore be incorporated into the NetzDG, which in turn extends the law's scope to video-sharing platforms (VSPs).[26]

[25] Due to the AVMSD being a European Union Directive, it is up to the individual member states to draft legislation that respects the obligations as set out in the European directives. Germany's adoption is covered in the April 2020 legislation.
[26] In 2018, the EU updated its Audio-Visual Media Services Directive (AVMSD), which governs Union-wide coordination of national legislation on audio-visual services (such as television broadcasts), to include online video-sharing platforms (VSPs). It encourages Member States to ensure that VSPs under their jurisdiction comply with the requirements set out in the AVMSD, including preventing the dissemination of terrorist content.

Main regulatory bodies:

- The Voluntary Self-Regulation Multimedia Service Providers (FSM) is a self-regulatory body recognised by the NetzDG. The review panel consists of 50 lawyers, and tech companies can appeal to the FSM when they are unsure of the illegality of content reported to them. Only social networks that are members of the FSM can do so.
- As a general rule, the German government conscripts tech companies in remit of the law to carry out the requirements of the legislation set by the German government.

Main takeaways for tech platforms:

- The *NetzDG* is one of the most extensive regulations of online content in the world. It requires tech companies to:
  - Introduce an "effective and transparent complaint mechanism" for users to swiftly report criminally liable (under the German Criminal Code) content
  - Assess reported content's illegality under German law and remove content quickly. Rules stipulate that once notified by users, a company shall remove "manifestly unlawful content" within 24 hours and other prohibited content within 7 days
  - Produce compulsory bi-annual transparency reports detailing how they respond to user reports
  - Pay fines of up to either 5 (for individual responsible for the complaints mechanism) or 50 million euros (for company itself) when failing to comply with the regulation
- The *April 2020 Bill* adds further requirements to the NetzDG by compelling companies to:
  - Improve the quality of their transparency reporting, requiring tech companies to:
    - Provide information on counter-notification procedures
    - Detail the results of their use of automated methods for detecting illegal content
    - Clarify whether they have given access to their data to independent researchers
  - Facilitate reporting processes of illegal content
  - Strengthen appeal processes to allow users to challenge content removal decisions through a case-by-case review process
- The April 2020 Bill also includes the February 2020 amendment, *Gesetzentwurf zur Bekämpfung des Rechtsextremismus und der Hasskriminalität,* which is currently on hold. This amendment would require companies to:
  - Provide the Federal German Police Force with private information of users posting illegal content
  - Prohibit tech companies from alerting users about the action taken for 14 days

October 2017: The NetzDG

The NetzDG was introduced in 2017 to combat hate speech and target terrorist and extremist content, misinformation, and online speech that "may lead to hate crimes". The NetzDG is aimed at large social media companies with over 2 million users.

When unveiled, the NetzDG was criticised by several civil society organisations, including Article 19 and Human Rights Watch, as well as by David Kaye, the United Nations Special Rapporteur on Freedom of Expression. Kaye criticised the fact that it is now the legal responsibility of tech companies to adjudicate on the illegality of content, with little to no accountability from courts and public prosecutors. According to Kaye, since the German Criminal Code and tech platforms' Terms of Service are different, tech companies are now responsible for following two sets of guidance on content moderation, without court orders or judicial review to assist them in determining content illegality.

Article 19, on their part, cautioned that the 24-hour removal deadline and high fines faced by platforms might make companies err on the side of removal. This could lead to the censoring of content that is neither extremist nor illegal in nature – what some have called "over-policing" of content – and poses serious questions with regards to potential negative impact on freedom of expression. Daphne Keller has pointed out that – whilst some argue that the NetzDG has not led to an increase in content removal – without confirmation of how companies have increased their Terms of Service removals as a precautionary action as a result of NetzDG, there is no way to adequately assess this.

February 2020: Gesetzentwurf zur Bekämpfung des Rechtsextremismus und der Hasskriminalität

Germany's parliament passed an amendment to the NetzDG in February this year. The amendment aims to further regulate hate speech, cyberbullying, and extremist content that stems from violent far-right extremism through obliging tech companies to share the information of users that post illegal content to the German Federal Police Force.

The amendment follows three right-wing terrorist attacks in Germany; the 2019 Halle attack, the 2020 Hanau attack, and the 2019 murder of pro-immigration politician Walter Luebcke. Luebcke's murder was highlighted by the German government when justifying the introduction of the amendment, stressing that his death as was preceded by him being targeted online hate speech.

The amendment underwent a review on 7 October 2020 and has been put on hold. German President Frank-Walter Steinmeier has held off from ratifying the amendment due to its potential unconstitutionality, mainly because of possible privacy violations. This mirrors civil society and legal concerns over the amendment, which consolidated around serious privacy concerns if social media platforms were to provide the government with users' private information, without any judiciary oversight. At the time of writing, it is unclear what will happen to the amendment, but content regulation expert Matthias Kettemann has hypothesised that the amendment might be brought before parliament again or repealed altogether.

April 2020: further amendments to the NetzDG

The 2020 April draft bill, which is a collation of further amendments to the NetzDG, widens the scope of the NetzDG from social media platforms to VSPs, extends the requirements put on tech companies (see above), and includes the above-mentioned February 2020 amendment.

EuroISPA, a pan European association of European Internet Services Providers Associations (ISPAs), has raised concern over the German amendments being drafted before the EU Digital Services Act has been updated, as the DSA was still undergoing public consultation when the 2020 April amendments were drafted. EuroISPA cautioned that the NetzDG amendment and its implication for online regulation in Germany will limit legislative consistency across member states, which in turn will affect tech companies and VSPs as they need to respect individual member states' online regulation rules, which limits new starters from entering the market. This is of particular concern in the April 2020 amendments, as this extends the scope of the NetzDG to VSPs of all sizes, as the amendment is meant to incorporate Germany's obligations under the EU's AVMSD (2018) into the NetzDG scheme. Whilst bigger companies might have the resources to comply with the NetzDG, smaller companies might struggle due to lack of capacity. Given that terrorists predominantly exploit smaller tech platforms for this very reason, this presents significant risks to competition and innovation.

Concerns over negative global impact

The NetzDG has been used as a template for regulatory frameworks in other countries, despite the significant critiques of the law. Several civil society groups have warned that the law may inspire similar or more restrictive regulation by less democratic nation states, which could further infringe on freedom of speech and digital rights globally.

Resources

Article19 (2017), *Germany: Act to Improve Enforcement of the Law in Social Networks.*

de Streel Alexandre et al (2020), *Online Platform's Moderation of Illegal Content Online, Policy Department for Economic, Scientific and Quality of Life Policies – Directorate-General for Internal Policies.*

Earp (2020), *Germany Revisits Influential Internet Law as Amendment Raises Privacy Implications,* Committee to Protect Journalists.

Echikson (2020), *The Impact of the German NetzDG Law*, CEPS Europe.

Kaye (2019), *Speech Police: The Global Struggle to Govern the Internet: Columbia Global Reports.*

Hardinghaus, Kimmich, and Schonhofen (2020), *German Government Introduces New Bill to Amend Germany's Hate Speech Act, Establishing New Requirements for Social Networks and Video-Sharing Platforms,* Technology Law Dispatch, ReedSmith.

Heldt (2020), *Germany is amending its Online Speech Act NetzDg... but Not Only That,* Internet Policy Review.

Human Rights Watch (2018), *Germany: Flawed Social Media Law.*

Lee (2017), *Germany's NetzDG and the Threat to Online Free Speech*, Yale Law School, Media, Freedom and Information Access Clinic.

Lomas (2020), *Germany Tightens Online Hate Speech Rules to Make Platforms Send Reports Straight to the Feds,* Techcrunch.

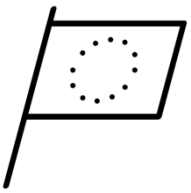
Pielemeier (2019*), NetzDG: A Key Test for the Regulation of Tech Companies*, Global Network Initiative .

Tworek and Leersen (2019*), An Analysis of Germany's NetzDG Law*, Transatlantic Working Group.

European Commission (2020), *Audiovisual Media Services Directive (AVMSD).*

Parliament of Germany (2017), *Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act).*

EUROPE | THE EUROPEAN UNION



The European Union (EU) is an influential voice in the global debate on regulation of online speech. For that reason, two upcoming regulatory regimes might – in addition to shaping EU digital policy – create global precedents for how to regulate both online speech generally and terrorist content specifically.

EU's Regulatory framework:

- European Counter Terrorism Strategy, adopted in November 2005, which sets out the EU's priorities on countering terrorism in the Union.
- European Agenda on Security, adopted in April 2015, which announced the establishment of key institutions to tackle terrorist use of the internet such as the EU Internet Referral Unit and the EU Internet Forum.
- Directive (EU) 2017/541 on combating terrorism, adopted in March 2017, and the key EU legal act on terrorism.[27]
- E-Commerce Directive, adopted in June 2000, which provides the overall framework for the EU's Digital Market and dictates that tech companies are exempt from liability for user-generated content.
- Audio Visual Media Services Directive, adopted in November 2018, which compels Member States to prevent audio-visual services, including online video-sharing platforms, from disseminating harmful material, including terrorist content.

Proposed regulation:

- *Regulation on preventing the dissemination of terrorist content online* (proposed by the European Commission in 2018 and currently in trilogue[28] process), which proposes to compel tech companies to remove terrorist content within one hour and introduce proactive measures to filter such material.[29]
- *Digital Services Act* (DSA), announced in 2020 as part of the new European Commission's aim for a "Europe fit for the Digital Age". A proposal for the DSA was announced on 15 December, making some changes to the pre-existing liability scheme and requiring companies to establish a notice and action mechanism. The EU opened a consultation process on the DSA which closed in September, and you can read our response here.

[27] In EU law-making, a "Directive" is a legislative act sets out goals that all EU countries must achieve, however without specifying exactly how to reach these targets. For more information, see: https://europa.eu/european-union/law/legal-acts_en
[28] The negotiation process between the EU's three legislative bodies: the European Commission (which proposes regulation), the EU Parliament, and the Council of the EU, who are able to suggest changes to the proposed text before its adoption.
[29] Unlike a Directive, a Regulation is legally binding and must be applied in its entirety across the EU.

Key organisations and forums:

- Europol, the European Union's law enforcement agency which supports Member States in countering organised crime and terrorism.
- EU Internet Referral Unit, (Europol), which reports terrorist content to tech platforms for their assessment and removal based on platform Terms of Service.
- EU Internet Forum, a public-private forum set up by the Commission to tackle terrorist use of the internet.

Collaborative schemes:

- EU Code of Conduct on Illegal Hate Speech, in which signatory tech companies commit to remove and report on hate speech flagged to them by a select number of European civil society groups.
- EU Crisis Protocol, a collaborative mechanism between governments and tech companies for the rapid detection and removal of terrorist content in the event of an online crisis.

Key takeaways for tech platforms:

- Companies are currently exempt from legal liability for user-generated content, although this could change as part of the new Digital Services Act.
- There is a possibility that removal deadlines and demands for proactive measures to tackle terrorist content will be introduced as part of new Union-wide regulation.
- Companies have the possibility to participate in a number of voluntary collaborative schemes together with European law enforcement agencies and Member States.
- The EU is an influential regulatory force, and there is reason to believe that EU regulation could inspire similar efforts elsewhere.

EU counterterrorism strategy

The EU's Counter Terrorism Strategy, launched in 2005, provides a framework for the Union to respond to terrorism across four strands: prevent, protect, pursue, and respond. Whilst the strategy does not focus on terrorist use of the internet, it does mention the need to counter this as part of its "prevent" strand.

Many of the texts and bodies involved in tackling terrorist use of the internet in the EU came into fruition around 2015. In April of 2015, the EU adopted the European Agenda on Security, which addresses preventing terrorism and radicalisation that leads to terrorism at length, including terrorist use of the internet. The Agenda also committed the EU to setting up two collaborative schemes: Europol's EU Internet Referral Unit (EU IRU) and the EU Internet Forum.

The key regulatory document guiding the EU-wide counterterrorism response is Directive 2017/451 (also known as the "Terrorism Directive"). The Directive replaced previous texts (such as Council Framework Decision 2002/475/JHA) and provides definitions of key terms, including of "terrorist groups," "terrorist offences", and terrorist propaganda ("public provocation to commit a terrorist

offence"). The Directive was partly introduced to better reflect the need to tackle terrorist use of the internet, and lays down guidelines for Member States to address this threat. For example, the Directive instructs Member States to ensure "prompt removal" of online terrorist content, whilst stressing that such efforts should be based on an "adequate level of legal certainty" and ensure that there are appropriate redress mechanisms in place.

Online terrorist content: current regulatory landscape

The main legal act outlining tech company responsibilities with regards to illegal and harmful content is the E-Commerce Directive of 2000. Whilst initially meant to break down obstacles to cross-border online services in the EU, the E-Commerce Directive also exempts tech companies from liability for illegal content (including terrorist content) that users create and share on their platforms, provided they act "expeditiously" to remove it.[30] Further, Article 15 outlines that tech companies providing have no obligation to monitor their platforms for illegal content. This arrangement is being reconsidered by the EU, both through the proposed regulation to combat online terrorist content and the Digital Services Act.

In 2018, the EU updated its Audio-Visual Media Services Directive (AVMSD), which governs Union-wide coordination of national legislation on audio-visual services (such as television broadcasts), to include online video-sharing platforms (VSPs). It encourages Member States to ensure that VSPs under their jurisdiction comply with the requirements set out in the AVMSD, including preventing the dissemination of terrorist content. In a communication, the European Commission specified that VSP status primarily concerns platforms who either have the sharing of user-generated video content as its main purpose or as one of its core purposes, meaning that in theory the AVMSD could apply to social media platforms on which videos are shared, including livestreaming functions.

Proposed regulation on preventing the dissemination of terrorist content online

In September 2018, the EU Commission introduced a proposed "regulation on preventing the dissemination of terrorist content online". The regulation has since undergone the EU's legislative trilogue process of negotiation between the Commission, Parliament, and the Council. To date, only Parliament's reading of the proposal has been published in full.

The proposal suggests three main instruments to regulate online terrorist content:

- Swift removals: companies would be obliged to remove content within one hour of having received a removal order from a "competent authority" (which each Member State will be able to appoint). Failure to meet the one-hour deadline could result in penalty fees of up 4% of the company's global annual turnover.
- Content referral: the competent authority will also be able to refer content to companies, similar to the role currently played by the EU IRU, for removal against company Terms of Service.

[30] This has some similarity to the US Section 230 of the US Communications Decency Act exempts tech companies from legal liability for user-generated content located on their platforms.

- Proactive measures: companies would be required to take "proactive measures" to prevent terrorist content from being uploaded on their platforms – for example by using automated tools.

The Commission's proposal drew criticism from academics, experts, and civil society groups. Further, the proposed regulation was criticised by three separate UN Special Rapporteurs, the Council of Europe, and the EU's own Fundamental Rights Agency, which said that the proposal is in possible violation of the EU Charter for Fundamental Rights. Criticism mainly concerns the short removal deadline and the proactive measures instrument, which according to critics will lead to companies erring on the side of removal to avoid penalty fees.

Whilst the regulation clarifies that its definition of "terrorist content" is based on the Terrorism Directive, there have been concerns that companies – due to the risk of fines – might remove content shared for journalistic and academic purposes. There has also been criticism raised against the referral mechanism, since this allows for tech company Terms of Service, as opposed to the rule of law, to dictate what content gets removed for counterterrorism purposes. Content moderation expert Daphne Keller has called this the "rule of ToS." At Tech Against Terrorism, we have cautioned against the proposal's potential negative impact on smaller tech companies, and warned against the potential fragmentation that it risks leading to. We also encourage the EU to provide more clarity as to what evidence base motivates the one-hour removal deadline.

The EU Parliament's reading of the proposal, unveiled in April 2019, provided some changes, for example by deleting the referral instrument and limiting the scope to "public" dissemination of terrorist content to avoid covering private communications and cloud infrastructure. These changes were largely welcomed by civil society groups. Although a version of the proposal worked on by the Council, which reintroduces some of the elements that Parliament modified, was leaked in March 2020, there has been no confirmation as to what a final version of the regulation will look like.

EU-led voluntary collaborative forums to tackle terrorist use of the internet

Whilst there is currently no EU-wide legislation regulating terrorist use of the internet, the EU has been influential in encouraging tech company action on terrorist content via a number of forums.

- EU Internet Forum (EUIF), bringing together Member States, tech companies, and relevant expert stakeholders (Tech Against Terrorism has participated in EUIF meetings since 2017) with the aim of creating joint voluntary approaches to preventing terrorist use of the internet and hate speech. Whilst there have been concrete outcomes of the Forum, such as the EU Code of Conduct on Hate Speech and the EU Crisis Protocol, voluntary arrangements like EUIF have been criticised for setting undue speech regulation under the guise of volunteerism. One notable critic is Professor Danielle Citron, who has described the EUIF as an example of the EU contributing to "censorship creep".[31] According to Citron, several of the voluntary steps that tech companies have taken to address terrorist use of their platforms

[31] By censorship creep, Citron means that online counterterrorism efforts or mechanisms risk taking on functions and having reach beyond its intended purpose, which risks leading to censorship of legal and legitimate speech online.

since 2015 have been made specifically to placate EU legislators. Whilst Citron acknowledges that results have come out of this approach (the GIFCT hash-sharing database is one example), the definitional uncertainty around terms like terrorist content means that there is significant risk of erroneous removal negatively impacting freedom of expression. Further, since companies are tackling content "voluntarily", material is removed under company speech policies rather than local or regional legislation, meaning that effects are global effects despite being based on European standards.

- EU Internet Referral Unit (EU IRU), based on the model pioneered by the UK's Counterterrorism Internet Referral Unit. The EU IRU employs subject matter experts to trawl the web and refer suspected Islamist terrorist content to tech companies, who then assess whether the content violates their Terms of Service. Member States are also able to refer content to the EU IRU. The unit conducts so-called referral assessment days with tech companies. This has led to substantial removal of terrorist content, including a joint operation with Telegram to remove a large number of Islamic State channels. According to the EU IRU, the Unit has to date referred more than 111,000 pieces of content to tech companies. Whilst this approach has been commended, criticism has been leveraged against the EU IRU (and IRUs generally) due to their risk of undermining rule of law by promoting content removal via extra-legal channels as content is removed based on company ToS rather than legal statutes. Whilst the Unit does release annual transparency reports, the Global Network Initiative (GNI) has noted that there is no formal oversight of judicial review of the EU IRU's activities.

Resources

AccessNow (2020), *How the Digital Services Act could hack Big Tech's human rights problem.*

Article19 (2020), *Article 19's Recommendations for the EU Digital Services Act.*

Citron (2018), *Extremist Speech, Compelled Conformity, and Censorship Creep*, Notre Dame Law Review

Europol (2019), *EU IRU 2018 transparency report.*

Europol (2020), *EU IRU 2019 transparency report.*

Kaye, Ni Aoilain. Cannataci (2018) *Letter from the mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights, UNOHCR.

Keller 2019), *The EU's terrorist content regulation: expanding the rule of platform terms of service and exporting expression restrictions from the EU's most conservative member states*, Stanford Cyber Policy Center.

Hadley & Berntsson (2020), *The EU's terrorist content regulation: concerns about effectiveness and impact on smaller tech platforms*, Vox-Pol.

Tech Against Terrorism (2020) *Summary of our response to the EU Digital Services Act consultation process.*

European Commission, *Proposal for A Regulation of The European Parliament and of The Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.*

European Commission (2020), *The Digital Services Act.*

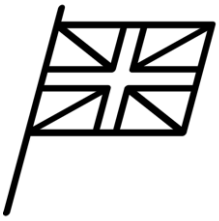
European Commission, *Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA.*

European Commission (2015), *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: The European Agenda on Security.*

European Commission (2018), *Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online.*

Council of the European Union (2005), *The European Union Counter-Terrorism Strategy.*

EUROPE | THE UNITED KINGDOM



The United Kingdom has set out an ambitious online regulatory framework in its Online Harms White Paper, aiming to make the UK "the safest place in the world to be online" by countering various online harms ranging from cyberbullying to terrorist content. This is yet to come into effect, but the UK has approved an interim regime to fulfil obligations under the European Union Directive, which the UK needs to comply with during Brexit negotiations. The UK also has extensive counterterrorism legislation criminalising the viewing and sharing of terrorist content online.

UK's regulatory framework:

- *The Online Harms White Paper* was published in April 2019 and outlines the key principles for online regulation in the UK. The paper suggests that tech companies should have a "mandatory duty of care" to protect users from "online harms". The draft law has since undergone consultation and is expected to be introduced into parliament in 2021.
- *The Terrorism Act 2000* is a cornerstone of UK terrorism legislation. Section 58 of the Act specifies the offence of possessing information, including via online means, that is "useful to a terrorist" .
- *The Terrorism Act 2006* creates new offences related to terrorism, as well as amends existing ones. A relevant example is Section 2, which makes it an offence to disseminate terrorist propaganda for "terrorist purposes".
- *The Counter-terrorism and Border Security Act 2019* amends section 58 of the Terrorism Act. It also criminalises obtaining or viewing such material online.
- *The Interim Approach*, put in place whilst awaiting introduction of the Online Harms regime, is due to come into effect on 1 November 2020. It sets out an interim regime for online VSPs to meet the UK's obligation of content regulation under the EU's Audiovisual Media Services Directive (AVMSD) 2018. The Government has transposed the VSP framework into Part 4B of the Communications Act 2003 ("the Act")

Main body overseeing online regulation:

- Ofcom, the UK communications regulatory body. Ofcom oversees new regulations, both under the Interim Approach and the proposed online harms legislation.

Key bodies and institutions:

- The UK Internet Referral Unit (CT IRU) detects and refers terrorist content to tech platforms for assessment against companies' Terms of Service.
- The Department for Digital, Culture, Media and Sport (DCMS)_is partly responsible for legislation relating to the Internet and media broadcasting. Together with the Home Office, the DCMS initiated the Online Harms White Paper.
- The Home Office is responsible for security and policing in the UK, including counterterrorism and terrorist use of the internet.
- The Independent Reviewer of Terrorism Legislation_scrutinises and reports on terrorism legislation in the UK. The current reviewer is Jonathan Hall.

Key takeaways for tech companies:

Interim Regime:

As the regulation is set to come into effect on November 1, Ofcom stated that it expects VSPs to assess whether they fall in the remit of the new legislation and to conduct risk assessments to identify what potential harms are to their users.

- When in remit of the law, VSPs, regardless of their size, need to protect users under the age of 18 from accessing restricted material[32].
- Regardless of their size, VSPs need to protect all users from "relevant harmful material":
    - Relevant harmful material constitutes "any material likely to incite violence or hatred against a group of persons or a member of a group of persons based on particular grounds";
    - "It also refers to the inclusion of any material which would be a criminal offence under laws relating to terrorism, child sexual exploitation or racism and xenophobia".
- In doing so, platforms need to regulate such content based on "proportionality"
    - In order to decide that proportionality, VSPs need to take into account the size and nature of the service, the type of harm caused, the exposed user's characteristics and the implications for freedom of expression
- VSPs need to implement an "out of court redress mechanism" to allow for user appeal of content that may have been removed erroneously.
- Ofcom can request VSPs to share information detailing the measures taken on different complaints.
- Ofcom can serve enforcement notices and financial penalties of up to £250,000 or 5% of the company's "qualifying revenue".
    - Ofcom has stated that in the "early regulatory period", it will only serve its enforcement mechanism in instances of a serious breach in compliance showcased by an absence of measures taken by VSPs.
    - However, it is unclear what will happen when this "early regulatory period" ends.

[32] Restricted material constitutes "videos which have or would be likely to have an R18 certificate, or which have been or would like be refused a certificate. It also means other material that might impair the physical, mental or moral development of persons under the age of 18".

Online Harms proposal:

- The proposed legislation will cover a wide range of "harmful content", including that which involves child sexual exploitation, cyber bullying, incitement to violence, encouragement of suicide, and terrorist and extremist content.
- A two-tier system will be imposed, with terrorist content and child sexual exploitation requiring more extensive action by tech companies than other harms.
- All tech platforms that permit online interactions and sharing of content will fall within the legal remit of the new mandatory duty of care to protect users from viewing harmful content online. The proposal suggests the following requirements for tech companies to uphold duty of care:
    - Update Terms of Service (ToS) to explicitly mention which content they deem appropriate (or inappropriate) on their platforms;
    - Produce annual transparency reports;
    - Introduce an easy-to-access user complaints function;
    - Respond to user complaints in an "appropriate timeframe" (to be set by Ofcom).[3]

- A "tiered enforcement system" will be implemented for companies that fail to uphold the "duty of care", escalating from:
    - Substantial fines (no amount has been specified yet);
    - Blocking of sites;
    - Criminal liability for members of a platform's senior management;[4]
    - Internet Service Provider (ISP) blocking for the most "severe" cases.

Terrorism Act 2000, the Terrorism Act 2006 & the Counterterrorism and Border Security Act 2019

In an amendment to article 58 of the Terrorism Act 2000, as written in the Counterterrorism and Border Security Act, viewing terrorist content online just once may give up to 15 years in prison. However, penalisation is dependent on knowing the purpose of that content (it being terrorist in nature), without a reasonable excuse (including journalistic or academic work).

The former Independent Reviewer of Terrorism Legislation, Max Hill, raised questions on the amendment on subsection 58 of the Terrorism Act when it was proposed in 2017. In a response, he and Professor Clive Walker of Leeds University School of Law asked whether an amendment was needed in the first place. They concluded that the existing clauses 1 (the encouragement of terrorism), 2 (the dissemination of terrorist publications), and 5 (the preparation of terrorist acts) of the Terrorism Act 2006 were sufficient for prosecuting and criminalising the online viewing of terrorist content, and so argued that the amendment was not necessary.

The Independent Reviewer subsequently considered the proposed amendment, which at that time still set out to criminalise "repeated viewing" of terrorist content on the Internet. On this premise, the Independent Reviewer identified that the law had the potential to "catch far too many people". However, as mentioned above the final Act went a step further, dropping the "repeated viewing" element and criminalising one-off viewing of terrorist material. The Independent Reviewers' concerns were publicly shared by civil society groups, who cautioned that it might have detrimental impact on freedom of speech.

The Independent Reviewer's original criticism also identified potential issues with users having to understand the "purpose of content" in order for the law to be effective, arguing that viewing of terrorist content does not necessarily mean that a user understands its purpose. This line of criticism can also be applied to sharing and disseminating content, as again, users might not be aware that the content is there for "terrorist purposes".

Furthermore, the United Nations special rapporteur on human rights and counter-terrorism, Professor Fionnuala Ní Aoláin, criticised the Counterterrorism and Border Security Act 2019 for being based on a "conveyer-belt" understanding of radicalisation or taking up violence, pointing out that there is little academic support for the theory that an individual will become radicalised by viewing terrorist content alone. Ní Aoláin also stated that whilst there are some protections for academics and journalists, other users will be infringed in their right to impart, seek, and receive information.

Online Harms White Paper

The Online Harms White Paper was published in April 2019 by the UK Home Office and the UK DCMS.

The proposed legislation has not yet entered parliament, but a Consultation process was held in 2019. In total, 2,400 responses were received from a broad range of stakeholders, including larger and smaller tech companies, governments, academics, think-tanks, civil society groups, and publishers.[5]

The White Paper covers a broad and varying range of online harms, although it distinguishes between "potentially harmful content" and "illegal content". Illegal content includes child sexual exploitation as well as terrorist content. This distinction was made to ensure the proportionality of the legislation, meaning that extreme content requires "further action" from platforms. However, the legislation does not define terrorist content or what going "further" entails. The proposal limits itself to suggesting that content removal should be preferred for illegal content, whilst other online harms should be addressed by other "content processes in place by tech companies".[33]

The proposed legislation has received criticism in the following areas:

- Human rights and rule of law concerns: civil society groups, law practices and tech initiatives have criticised the lack of clarity on what constitutes "online harms". This would leave tech companies with the responsibility of adjudicating what constitutes illegal content and what classifies as "potentially harmful content", without guidelines on how to assess such material. Due to the enforcement mechanism that awaits tech companies if they fail to identify and address content effectively (high fines as well as potential liability), civil society groups such as Article 19 have warned that this may incentivise companies to err on the side of content removal for both potentially illegal and "harmful" content. This risks the removal of legal and innocuous content, thus hindering digital rights, particularly freedom of speech. It also risks

[33] UK Consultation Report.

labelling content in the online space as "potentially harmful" or even illegal, despite it being legal offline. Finally, the Global Network Initiative has warned against imposing liability on tech companies on the basis that it likely to lead to the over-removal of content rather than tackling the underlying drivers of terrorist content on the Internet.

- Competition, capacity and innovation concerns: tech initiatives such as Coadec and TechUK have highlighted that smaller tech companies might not have the ability or resources to comply with the proposed requirements, which they say risks harming competition and innovation.

- *Legal concerns:* Legal experts have questioned the legality of imposing potential intermediary liability on managers at tech platforms, especially how the Online Harms legislation will work alongside the E-Commerce Directive 2000, which protects tech companies from liability. Legal critics also raised concern over the steep fines and the consequences that might lead to the over removal of content. In addition, Article 15 of the same Directive stipulates that Member States cannot impose general monitoring obligations for Internet platforms, which also raises questions on the extent to which the proposed legislation will uphold this Directive.[34] Civil society groups have added that the UK Communications Act, which ensures the protection of freedom of speech, risks being undermined by the proposed legislation.

The Interim Regime

The Interim Regime will work to ensure that the UK upholds its obligations under the EU's AVMSD until the Online Harms legislation is passed. As such, the Interim Regime applies to all UK VSPs. The EU updated the AVMSD, which governs Union-wide coordination of national legislation on audio-visual services (such as television broadcasts), in 2018 to include VSPs. It encourages Member States to ensure that VSPs operating under their jurisdiction comply with the requirements set out in the AVMSD, including preventing the dissemination of terrorist content. The European Commission has specified that VSP status primarily concerns platforms who either have the sharing of user-generated video content as its main purpose or as one of its core purposes, meaning that in theory the AVMSD could apply to social media platforms as well.

Similar to the feedback raised on the *Online Harms White Paper*, criticism raised by legal experts, civil society groups, and tech companies on the Interim Regime consolidate around the enforcement mechanisms that might lead to over-removal and potentially hinder competition and innovation as well as the lack of definitional clarity when it comes to defining harmful content, and particularly terrorist content.

However, Ofcom's most recent guidance for VSPs specifies that its first priority is to work together with the VSPs to strengthen or implement new measures in order to comply with the interim regime in its "early regulatory phase". In addition, Ofcom has provided guidance on how to determine proportionality between the action taken by a VSP and the level of harm of a particular piece of

[34] The UK only has existing obligations to the European directives for the duration of the Brexit negotiations; therefore, the legal concerns might become less relevant. However, whilst the UK might not have to fulfil the European directives, potential implications for freedom of speech and intermediary liability are still valid for post-Brexit Britain.

content. Ofcom stipulates that the size of the VSPs will be taken into account in both its proof of compliance as well as Ofcom's enforcement mechanism. Whilst this guidance clarifies some of the new requirements put on VSPs, the guidance is likely to change throughout the early regulatory phase.

Tech Against Terrorism offered a response to Ofcom's consultation process on the regulation of VSPs, which was concluded in September, which can be found here.

Resources

Article19, (2019) *Response to the Consultations on the White Paper on Online Harms.*

Global Network Initiative (2020), *Content Regulation and Human Rights.*

Human Rights Watch (2020), *Social Media Platforms Remove Evidence War Crimes.*

Lomas (2019), *UK Sets Out Safety-focused Plan to Regulate Internet Firms*, Techcrunch.

Osborne (2020), *Online Harms Regulation | Clarity Awaited but Reforms Set to Be Delayed.*

Tech Against Terrorism (2020), *Summary Tech Against Terrorism's Response to Ofcom's Consultation Process on the Regulation of Video-Sharing Platforms.*

Government of the United Kingdom (2020), *Online Harms White Paper - Initial Consultation Response*.

Vinous, (2019) *TechUK comments on the Government's new Online Harms White Paper*, TechUK.

Ofcom (2020), *Regulating video-sharing platforms A guide to the new requirements on VSPs and Ofcom's approach to regulation.*

Government of the United Kingdom, *Counter-Terrorism and Border Security Act 2019.*

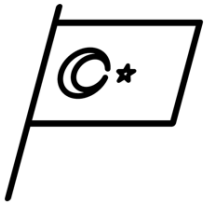
Government of the United Kingdom, *Terrorism Act 2000.*

Government of the United Kingdom, *Terrorism Act 2006.*

Government of the United Kingdom (2019), *Online Harms White Paper*

.

EUROPE | TURKEY



Online content regulation in Turkey is characterised by extensive removal of material that has resulted in a large number of Turkish and international websites being blocked in recent years. Further, the Turkish government recently introduced a Social Media Bill, implementing a wide range of new regulations and steep penalties for social media companies, which critics say poses further threats to online freedom of expression in the country.

Regulatory framework:

- *Bill Amending the Supreme Board of Radio and Television and Press Code, Law No. 4676,* May 2020, subjected the online space to restrictive press legislation in Turkey.[35]
- *The Regulation of Publications on the Internet and Suppression of Crimes Committed by means of Such Publication, 2007,* widely known as the "Internet Law 5651" or "Law No. 5651." This regulates prohibited content, such as child abuse images and obscenity, on the internet and enables the blocking of websites.
- Many provisions of the Criminal Code and other laws, such as Turkey's Anti-Terrorism Law and Defamation Law apply to online and offline activity. For instance, the Anti-Terrorism Law subjects those who "make online propaganda of a terrorist organisation" — "by legitimising, glorifying, or inciting violent methods or threats" — to imprisonment.
- *The Social Media Bill, Law No. 7253*, October 2020, compels social media companies with over a million daily users in Turkey to adhere to new regulations, such as storing user data in Turkey and a smaller timeframe for responding to complaints about posts violating personal and privacy rights, as well as fines for failure to comply.

[35] Apart from a single reported case, this law was never used by either prosecutors or courts.

Relevant national bodies:

- The Ministry of Transport, Maritime Affairs and Communications (MIT) is responsible for policy making for telecommunications in Turkey. Through its surveillance powers, the MIT is able to intercept and store private data on "external intelligence, national defense, terrorism, international crimes, and cybersecurity passing through telecommunications channels", without a requirement to obtain a court order.
- The Telecommunications Communication Presidency (TIB) was empowered in 2007 via the Internet Law 5651 to issue blocking orders for websites.
    - The TIB was shut down after the 2016 coup attempt in Turkey, and all of its responsibilities were transferred to the Information and Communication Technologies Authority (BTK).
- The BTK is an independent institution and has the power to enact by-laws, communications and other secondary regulations pertaining to the authorisations granted by the Electronic Communications Law.
- The Radio and Television Supreme Council (RTÜK), enabled by a March 2018 Bill, is authorised to regulate online content, including commercial streaming and foreign-based online media platforms.

Key takeaways for tech companies:

- The Internet Law 5651, or Law No. 5651, regulates the Internet and online service providers. Under this law:
    - ISPs are required to consolidate into a single "Association of Access Providers". Access providers part of the Association obtain an "activity certificate" to legally operate in Turkey, while those who are not members are not be able to provide services within the country.
    - Blocking orders can be issued by courts, public prosecutors, or the BTK.
        - Websites hosted in Turkey found to host proscribed content can be taken down, while websites based abroad can be blocked and filtered through ISPs.
        - Blocking orders can be administered if any individual or legal entity alleges a privacy violation, or if the content is considered "discriminatory or insulting to certain members of society". ISPs also have to block access to specific URLs within 4 hours of receiving an order.
        - Foreign-hosted websites are subject to blocking if they are suspected to contain eight categories of prohibited content, including: child abuse images, content that facilitates drug use, provision of substances dangerous to health, obscenity, prostitution sites, gambling sites, encouragement of suicide, and crimes committed against Mustafa Kemal Ataturk.
    - There are steep fines for failing to comply with the mentioned regulations:
        - If ISPs fail to comply with blocking orders within 4 hours, they face a fine up to 300,000 Turkish liras ($52,150), and if they fail to take action to block all alternative means of accessing the targeted site, such as proxy sites, it could result in a fine of up to 50,000 Turkish liras ($8,690).

- Under the new Social Media Bill, social media companies with over a million daily users in Turkey are required to:
  - Establish a formal presence in the country;
  - Respond to complaints about posts that "violate personal and privacy rights" within 48 hours, or face fines up to $700,000.
  - International companies are required to store user data in Turkey.
  - It would also allow courts to order Turkish news websites to remove content within 24 hours.
    - If social media companies do not comply with the new criteria within six months of the legislation having gone into effect, Turkish authorities will be able to ban advertising on the platforms, assign high fines, and adjust the sites' bandwidth by up to 90%.

- The RTÜK can regulate online content, including commercial streaming as well as foreign-based online media platforms. The RTÜK can also issue licenses to online content providers for a fee of 100,00 Turkish liras ($17,380) and is able to fine providers or revoke their licenses.
- Under Law No. 6532 on Amending the Law on State Intelligence Services and the National Intelligence Organisation (2014), the powers of the MIT to conduct surveillance were expanded and intelligence agents were granted unrestricted access to communications data without a court order.
  - Law No. 6532 mandates public and private bodies, such as banks, archives, professional organisations, and private companies, to provide the MIT with any requested data, documents, or information pertaining to certain crimes related to national security, state secrets, and espionage. Failure to comply can be punished with imprisonment.
  - Hosting and access providers must preserve all traffic information for one year and, in addition, access providers are required to provide assistance to the TIB (since 2016, the BTK) in monitoring internet traffic.

Anti-Terrorism Law and defamation offenses

According to the Freedom House 2020 assessment on Turkey, there are no laws that specifically criminalise online activities. However, many provisions of the criminal code and other laws, such as the Anti-Terrorism Law, are applied to online as well as offline activity. Article 7 of Turkey's Anti-Terrorism Law states, "those who make propaganda of a terrorist organisation by legitimising, glorifying, or inciting violent methods or threats" can be imprisoned for one to five years. This law has been criticised for its broad definition of terrorism, which has allegedly been exploited by courts to prosecute journalists and academics who criticise the government with no clear links to terrorist activities.

Online content regulation

In May 2002, the Turkish Parliament passed the *Bill Amending the Supreme Board of Radio and Television and Press Code* (Law No. 4676), which contained provisions that would subject the internet to restrictive press legislation in Turkey.

In May 2007, the government enacted the "Internet Law 5651" to regulate Internet and online service providers. The law regulates the liability of Internet intermediaries, including content, access, hosting service providers and "mass use" providers, such as "internet cafes", in Articles 4, 5 , 6, and 7. It therefore defines the responsibilities of content providers, hosting companies, public access providers, and ISPs. Although the law was introduced to hinder the spread of harmful content online, according to Human Rights Watch (HRW) it has been used to block LGBT community forums, independent media websites, and news sites with a pro-Kurdish political line. In addition, global websites hosting large volumes of user-generated content, including YouTube, Twitter, Blogspot, Wordpress, Vimeo, and Google Groups, have been blocked entirely on occasion, even if "only a fraction of the content was deemed subject to blocking". For example, following a corruption scandal that erupted when multiple conversations from top officials were leaked, authorities blocked access to Twitter and YouTube.

In February 2014, the law was amended, broadening powers to block content. This included requiring ISPs to consolidate into a single association as well as to block access to specific URLS within 4 hours of receiving an order. The amendments also expanded the TIB's powers to issue administrative blocking orders if any individual or legal entity alleges a privacy violation or if the content is considered "discriminatory or insulting to certain members of society". Non-compliance to these regulations result in high fines. These regulations were met with criticism, such as from HRW, who stressed that "although such blocking orders must be reviewed by a court within 48 hours, the grounds are so broadly and vaguely defined that they allow discretion for abusive application and interpretation". A typical concern is that high fines coupled with broadly and vaguely defined grounds might lead to companies erring on the side of overly-cautious content removal or blocking, which could lead to heightened censorship and infringe upon freedom of expression online.

According to HRW, as a consequence of the No. 5651 law, Turkish authorities have blocked tens of thousands of Turkish and international websites over the last few years.

New Social Media Bill

On October 1, 2020, Turkey's newest legislation for social media, "Law regarding the regulation of publications made in the Internet environment and the fight against crimes committed through these publications" came into effect, amending Law No. 5651. The law introduces strict regulations for social media companies, coupled with steep fines. This has been criticised by many human rights advocates for having potential negative effects on freedom of expression. According to HRW, this new social media law "paves the way for greater online censorship". HRW further argues that "Turkey's courts and regulatory bodies lack the independence necessary to prevent abuse of the law", and that "in practice the law therefore could serve as a new tool to silence critics online". The Global Network Initiative (GNI) has expressed concern as well, stating that this law "contains several

problematic provisions that are likely to significantly complicate the operation of social media services in Turkey and challenge the ability of Turkish citizens to freely exercise their rights to freedom of expression and privacy".

Regulation in practice: extensive online blocking

According to Freedom House's 2020 Freedom on the Net assessment, the Turkish constitution and laws "fail to protect freedom of expression and press freedom online", as online journalists and users frequently suffer civil and criminal penalties for legitimate expression. The state of emergency enacted in the aftermath of the 2016 coup attempt, which lasted until July 2018, allowed President Erdoğan to publish decrees without judicial oversight. This included decrees that were used to block websites, shut down communication networks, and close civil society organisations as well as news outlets. In particular, Decree No. 671 (2016) amended Turkey's Law on Digital Communications to empower the government to take "any necessary measure" on the grounds of "national security, public order, prevention of crime, protection of public health and public morals, or protection of the rights and freedoms" established under Article 22 of the Turkish constitution. It also requires any company that provides digital communications to enforce government orders within two hours of receiving them. Even though the state of emergency has not been in effect since 2018, the decree remains.

Resources

Akdeniz, *Report of the OSCE Representative on Freedom of the Media on Turkey and Internet Censorship*, OSCE.

Beceni Yasin, Sevim, Aslan. Zengin and Can Akdere (2017) *Communications: regulation and outsourcing in Turkey: overview*, Practical Law.

Freedom House (2020), *Freedom on the Net 2020: Turkey.*

Global Network Initiative (2020), *Content Regulation And Human Rights.*

Global Network Initiative (2020), *GNI Statement on Proposed Social Media Bill in Turkey.*

Human Rights Watch (2014), *Turkey: Internet Freedom, Rights in Sharp Decline.*

Human Rights Watch (2020), *Turkey: Press Freedom Under Attack.*

McKernan (2020), '*It's a war on words': Turks fear new law to muzzle social media giants,* The Guardian.

The Center for Internet and Society: Stanford Law School (2007), *Law No. 5651, May 23, 2007, Regulation of Publications on the Internet and Suppression of Crimes Committed by means of Such Publications.*

Zeldin (2014), *Turkey: Law on Internet Publications Amended*, Library of Congress.

MENA & SUB-SAHARAN AFRICA | KENYA

Kenya has "increasingly sought to remove online content", both through requests and increased regulation, that it deems "immoral" or "defamatory". Following terrorist attacks on civilian targets in recent years, the country has heightened its efforts around counterterrorism as well as online content regulation. Many of Kenya's legislations have been criticised by civil society for their "broadness", "vagueness", and potential "detrimental implications for freedom of expression". A proposed social media bill, if enacted, could largely impact social media companies and their users in Kenya, such as through strict regulations on user content.

Regulatory framework:

- *Kenya Information and Communications Act,* (KICA), October 1998, the primary legislation governing the telecommunications sector in Kenya. It has received numerous amendments since it first came into effect.
- *The proposed Kenya Information and Communication (Amendment) Bill*, 2019, also known as the "Social Media Bill", which would introduce stringent regulations on the use of social media in Kenya.
- *The Computer Misuse and Cybercrimes Act,* 2018, which establishes various offenses, including cyber terrorism, false publication of data, cyber harassment, identity theft and impersonation, and computer fraud.
- *National Cohesion and Integration Act,* 2008, which penalises hate speech and holds any media enterprise liable for publishing any utterance which amounts to hate speech.
- *Prevention of Terrorism Act* (PTA), 2012, Kenya's legal framework to combat terrorism.
    - Establishes terrorism related offenses and provides for special investigative powers, as well as special powers of arrest and detention of suspects.
- *Security Laws Amendment Act*, 2014, amended the PTA to strengthen the country's counter-terrorism efforts, and includes provisions on radicalisation and publishing offensive material.

Relevant national bodies:

- The Communications Authority (CA) is the regulatory authority for the communications sector in Kenya, established in 1999 by the Kenya Information and Communications Act. The CA is responsible for facilitating the development of the information and communications sectors, including broadcasting, cybersecurity, multimedia, telecommunications, electronic commerce, postal and courier services.
- The National Cohesion and Integration Commission (NCIC) is a statutory body that works to reduce interethnic conflict. It worked with the CA on the Guidelines to combat online abuse.

Key takeaways for tech companies:

- The Prevention of Terrorism Act and Security Laws Amendment Act enable national security bodies to intercept communications "for the purposes of detecting, deterring, and disrupting terrorism". They also include provisions on radicalisation as well as on the "publication of offending material".
- Guidelines implemented by the CA are set up to curb online abuse:
    - Prohibits political messages of specific violations;
    - Requires administrators of social media pages to moderate and control the content and discussions generated on their platform;
    - Gives mobile service providers the authorisation to block the transmission of political messages that, under their discretion, do not adhere to the CA's guidelines.
- The National Cohesion and Integration Act penalising hate speech can be invoked to remove or block online content, and has also been implemented in proactive action by service providers and other state agencies, such as the National Cohesion and Integration Commission (NCIC), in monitoring hate speech.
- The proposed Social Media Bill seeks to amend the KICA by introducing stringent regulations on the use of social media in Kenya, such as on the regulation of bloggers and social media platforms, and to introduce obligations for social media users.
    - The regulations include new requirements for the operators of social media platforms accessible in Kenya to obtain licenses and establish a physical office in the country. It further aims to place regulations on the content published by social media users.
    - It was tabled in parliament in Oct 2019, but the Bill has not progressed at the time of writing. Tech companies should keep an eye out for any developments on the Bill.

Prevention of Terrorism Act and Security Laws Amendment Act

In October 2012, the Prevention of Terrorism Act (PTA) was enacted to provide a "comprehensive and effective legal framework to combat terrorism" in Kenya. The Act allows authorities to limit constitutional freedoms during investigations into terrorism, such as the right to privacy. According to a Freedom House policy brief, the PTA only provides a "vague definition of terrorism, greatly expands, police powers, and allows the state to create lists of suspected terrorists and terrorist organizations without due process" and the "pervasive powers granted by the law have been to take abusive actions against organizations in contravention of constitutional requirements".

Some provisions of the PTA have since been amended by the Security Laws (Amendment) Act (2014) to strengthen the country's counter-terrorism efforts. This was proposed following a wave of terrorist attacks which heightened public pressure to curb those attacks. The law's progression was thus fast-tracked and enacted within ten days of its initial proposal. It amends 21 different laws, including the Penal Code, Criminal Procedure Code, Evidence Act, Prevention of Terrorism Act, and the National Police Service Act.

This Act enabled national security bodies to intercept communications "for the purposes of detecting, deterring, and disrupting terrorism", via an interception order from the High Court. It also introduces, among others, a new provision on radicalisation to the PTA. In doing so, it criminalises the adoption or promotion of "an extreme belief system for the purpose of facilitating ideologically based violence to advance political, religious or social change." A person convicted of this offense is subject to up to 30 years in prison. This provision has been criticised by rights groups, such as Human Rights Watch (HRW), who say that "the unclear language could be interpreted to prosecute political and human rights activists".

The amendment further implements a provision on the "publication of offending material". Under this section of the bill, anyone who "publishes or utters a statement that is likely to be understood as directly or indirectly encouraging or inducing another person to commit or prepare to commit an act of terrorism," is punishable by up to 14 years in prison. There have been concerns that this overly broad provision could be interpreted to apply to social media or any other public forum.

In December of 2014, the Coalition for Reforms and Democracy filed a suit in the High Court challenging the constitutionality of the law. The Kenya National Commission on Human Rights and other groups joined in the suit, challenging the law on the grounds that its provisions violated rights enshrined in the Constitution and that its passage violated parliamentary procedure, such as not involving the Senate. The Court dismissed the petition for the immediate suspension of the law.

Hate speech and Internet Intermediary Liability

Hate speech is penalised under the National Cohesion and Integration Act, which was enacted in response to widespread ethnic violence after the 2007 general elections. Under this Act, individuals found guilty of spreading hate speech can either face a fine of up to 1 million shillings ($9,600) or imprisonment of up to three years, or both.

Under the same National Cohesion and Integration Act (2008), another provision, s.62, holds any media enterprise liable for publishing any utterance which amounts to hate speech. A media enterprise can be fined up to 1 million shillings ($9,600) for publishing hate speech, which is broadly defined in the legislation. This provision can be invoked to remove or block online content and has also been implemented in proactive action by service providers and other state agencies in monitoring hate speech. Thus, Internet intermediaries in Kenya can be held accountable for illegal content, such as copyright infringements and hate speech. However, they are not required to actively monitor traffic passing through their networks unless they are made aware of illegal content.

Online content regulation

Prior to the 2017 election, the CA implemented new guidelines, to curb online abuse. This was a joint effort with the National Cohesion and Integration Commission (NCIC). The guidelines prohibit political messages that "contain offensive, abusive, insulting, misleading, confusing, obscene, or profane language".

These guidelines have been criticised for being very broadly worded, and due to their possible use to limit legitimate online expression. The guidelines additionally require administrators of social media pages to "moderate and control the content and discussions generated on their platform", for bulk political messages to require prior approval from the NCIC, and give mobile service providers the authorisation to block the transmission of political messages, via SMS and social media platforms, that, under their discretion, do not adhere to the guidelines.

The Kenya Information and Communication (Amendment) Bill, also known as the "Social Media Bill" was introduced in parliament in October 2019. The Bill seeks to amend the KICA by introducing stringent regulations on the use of social media in Kenya, such as on the regulation of bloggers and social media platforms, and to introduce social media user obligations.

The Bill proposes to require bloggers to obtain licenses from the CA. It has been criticised that the definition of blogging is wide and ambiguous, as it includes collecting, writing, editing and presenting of news or news articles on social media platforms. The bill defines social media platforms as including "online publishing and discussion, media sharing, blogging, social networking, document and data sharing repositories, social media applications, social bookmarking and widgets". There is concern that the definition is thus broad enough to include ordinary users of social media platforms such as YouTube, Facebook and Twitter.

It also proposes to require the operator of a social media platform which is accessible in Kenya to obtain a social media license as well as establish a physical office in the country. Here again, the Bill has been criticised for its broad definition, this time of a social media platform, which could include any online medium that allows for social networking and media sharing. It further aims to place a number of obligations on social media users, such as for them to ensure that their content is, among other things, accurate and unbiased, "does not degrade or intimidate a recipient of the content", and "is not prejudicial against a person or group of people based on their race, gender, ethnicity, nationality, religion, political affiliation, language, ability or appearance". If enacted in its current form, the Bill could have far reaching implications on the use of social media in Kenya. However, at the time of writing, the bill has not progressed.

Computer Misuse and Cybercrimes Act

In May of 2018, the Computer Misuse and Cybercrimes Act was introduced, establishing various offenses including, among others, false publication of data, cyber harassment, cyber terrorism, identity theft and impersonation, and computer fraud.

According to the Act, anyone who "knowingly publishes information that is false in print, broadcast, data or over a computer system, that is calculated or results in panic, chaos, or violence among citizens of the Republic, or which is likely to discredit the reputation of a person" can be subject to a fine of up to 5 million shillings ($48,000) and up to 10 years of imprisonment. The Act also requires service providers to assist in the investigation of offenses, such as by collecting and providing data to investigation officers. It also prescribes penalties for not complying with the provisions it sets in place, such as high fines and imprisonment.

Prior to the commencement of the Act, the Bloggers Association of Kenya (BAKE), with support from the rights group Article 19, filed a petition — on the basis that the law was unconstitutional and infringes on and threatens freedom of expression and the right to privacy, property and a fair hearing — that led a court to temporarily suspend 26 sections of the Computer Misuse and Cybercrimes Act before it came into effect in May 2018. However, in February of this year, the High Court dismissed the petition and lifted the suspension of the 26 sections, which set high fines and prison sentences for a variety of online activities, including publishing false information and cyber harassment. All of the provisions of the Act are therefore in full force and effect.

Recently, misinformation about the COVID-19 pandemic was widespread in the country. In response, the government threatened fines and imprisonment of up to two years, as charged under the Computer Misuse and Cybercrimes Act 2018, for people who spread false COVID-19 information online

Resources

Africa Times (2020), *Kenya warns of up to $50K fines for spreading fake COVID19 news.*

DLA Piper (2019). *Telecommunications Laws of the World: Kenya.*

Freedom House (2020), *Freedom on the Net 2020: Kenya.*

Freedom House (2018), *Online Survey: Kenya's Antiterrorism Strategy Should Prioritize Human Rights, Rule of Law.*

Hanibal (2014), *Kenya: Security Laws (Amendment) Bill Enacted*, Library of Congress.

Human Rights Watch (2014), *Kenya: Security Bill Tramples Basic Rights.*

Indokhomi and Syekei (2020), *The Computer Misuse And Cybercrimes Act,* Bowmans.

Kakah (2020), *Court dismisses bloggers' cybercrime law case,* Nation.

Munyua, Githaiga, and Kapiyo, *Intermediary Liability in Kenya*, Association for Progressive Communications (APC).

Mwathe and Syekei (2019) *Highlights on Proposed Law Introducing Strict Regulation of Social Media*, Bowmans.

The New Humanitarian (2012), *Analysis: Taming hate speech in Kenya.*

United Nations Office on Drugs and Crime (2017), Kenya Training Manual on Human Rights and Criminal Justice Responses to Terrorism.

Republic of Kenya (2014), *Kenya Gazette Supplement No. 167 (Acts No. 19).*

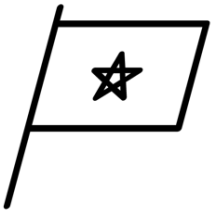
Republic of Kenya (2018), *Kenya Gazette Supplement – The Computer Misuse and Cybercrimes Act.*

Republic of Kenya (2019), *Kenya Gazette Supplement – The Kenya Information and Communications (Amendment) Bill.*

Republic of Kenya (2012), *Prevention of Terrorism Act.*

Republic of Kenya (2008), *National Cohesion and Integration Act.*

MENA AND SUB-SAHARAN AFRICA | MOROCCO



Morocco's online regulatory framework consists of different laws and codes that strive to limit the spread of content that can pose a threat to the Kingdom's "integrity, security and public order". Central to this framework are the 2003 Anti-Terrorism Law passed in the aftermath of the 2003 Casablanca bombings and the 2016 Press Code that lays out limitations journalistic publications and public speech. However, the existing regulatory framework is not explicitly clear regarding implications for tech platforms and the government's powers to filter the online space – something which has been criticised by civil society. According to Freedom House, the government also resorts to "extralegal means" to remove content that it deems "controversial or undesirable" by pressuring media outlets and online figures to delete such content.

Morocco's regulatory framework:

- *Loi n° 03-03 relative à la lutte contre le terrorisme,* May 2003, Morocco's legal framework for countering terrorism, providing definitions of key terms (such as "terrorist acts") and laying out the different sanctions and legal processes:
    - The law prohibits the diffusion of terrorist content by any means of speech (oral or written), including audio-visual and electronic material.
    - Beside acts of terrorism, the 2003 law also covers incitement to and condoning or apology of terrorism ("apologie du terrorisme"), as well as providing assistance to the preparation of a terrorist act and the non-disclosing of a terrorist offence.
    - Article 218.2 sanctions apology of terrorism with jail terms and fines, whilst article 218.5 sanctions incitement and "provocation" of terrorism.
- *Loi relative a la presse et a l'édition,* August 2016, Morocco's Press Code, regulating the press and public speech in general, including speech and journalistic content posted online:
    - Title III of the Press Code, related to the "special protection of certain rights",[36] specifies limitation to freedom of the press and public speech. In particular, it prohibits publication that threatens "public order", including those that prejudice "Islamic religion, the monarchy, or the integrity of the Kingdom".
    - Article 72, in particular, penalises the diffusion, by any means (including electronic) and by all individuals of:
        - Diffusion in bad faith of allegations, and of false or falsified information that have led to disruption to the public order or fear amongst population.[37]
        - Terrorism apology.
        - Incitement to hatred and racial hatred.

[36] Original title: "Des sanctions de la protection spéciale de certains droits. De la competence des jurisdictions et des procedures."
[37] Original provision: "quiconque a publié, diffusé ou transmis, de mauvaise foi, une nouvelle fausse, des allégations, des faits inexacts, des pieces fabriquées ou falsifiées attribuées à des tiers, lorsque ses actes auront troublé l'ordre public ou suscité la frayeur parmi la population et ce, quel que soit le moyen utilisé notamment par discours, cris ou menaces proférés dans les lieux ou reunions publics, par des écrits, des imprimés vendus, distribués, mis en vente ou exposés dans les lieux ou réunions publics"

    - Following articles under Title III further lays out restrictions to journalistic content and public speech, including related to the prohibition of defamation, libel, and slander.
    - Speech offences under the Press Code are sanctioned with various fines depending on the offence.
- *Code Penal Marocain (2018 consolidated version),* similarly to the Press Code the Kingdom's Penal Code sanctions certain non-violent speeches, including speech that are "showing a lack of due respect for the king, defaming state institutions, and insulting public agents while they are performing their duties".
    - However, unlike the Press Code, the Penal Code punishes speech offences with prison terms.
- *Draft law no. 22.20,* so-called "social media law", passed in March 2020 and later temporarily suspended in May 2020 whilst awaiting the end of the Covid-19 pandemic:
    - The draft law would task "network providers" with restricting access to and suppressing online content that pose a threat to security and public order within 24h.

Key takeaways for tech companies:

- Under the current legislative framework, internet platforms are exempt from liability for user-generated content, including terrorist content, with liability lying with the content's creator or poster.
- Article 37 of the Press Code stipulates that judicial authorities can request the (provisional) removal of online content that violates the dispositions specified under Title III of the same Code.

Anti-terrorism law and the penalisation of incitement and apology

Morocco's counter-terrorism framework was developed after the 2003 Casablanca bombings, which brought the Moroccan parliament to unanimously adopt the Anti-Terrorism law that had been debated since 2002. The night of 16 May 2003, Casablanca – Morocco's biggest city and economic hub – was hit by one of the worst terrorist attack in the Kingdom's history: 14 terrorists, all Moroccan citizens, launched a series of attacks on Belgian, Jewish and Spanish buildings, killing 45 people. The attack was coordinated by Salafiya Al Jihadiya, a terrorist group affiliated with al-Qaida. Labelled the "Moroccan 9/11", the attacks led to a strong counter-terrorist response from the Moroccan government, with the Anti-Terrorism law passed less than 10 days after, and over 200 people arrested in connection with the attacks.

The Anti-Terrorism law broadly defines terrorism undertakings that aims to "seriously undermine public order through intimidation, terror or violence", before specifying different acts that are considered "terrorist". Under the law, terrorist use of the internet is mostly addressed through the question of incitement and apology, both of which are penalised. The law strictly prohibits the diffusion by any means – whether by oral proclamation on the street or by audio-visual content shared by an electronic means – of speech that either condones or incites to acts of terrorism.

Individuals found condoning an act of terrorism or provoking others to commit such acts can face jail terms and substantial fines.

The Anti-Terrorism law does not directly address the responsibility and liability of tech platforms with regard to terrorist content online. There is also some uncertainty with regards to the scope of the powers conferred to Moroccan authorities regarding online content and terrorist exploitation of internet technologies. The law permits judicial authorities to request the interception and the seizure of communications in relation to a terrorist investigation, or in "extreme emergency" situations. It also sanctions with prison time, for individuals, or with a fine, for a legal entity, the non-disclosure of terrorist offences. However, the law in itself remains broad into specifying what the seizure of communications and non-disclosure entails. Potentially, both provisions can apply to internet service providers (ISPs). Freedom House's *Freedom on the Net* 2020 report on Morocco notes that "intermediaries must block or delete infringing content when made aware of it or upon receipt of a court order", and that the prosecution of complicity with an act of terrorism, specified in Article 218.6 of the Anti-Terrorism law, could potentially apply to site owners and (ISPs).

Regulation of public speech

Regulation of online content in Morocco is tied to restrictions specified in the 2016 Press Code, which limits journalistic content and online public speech by invoking the "special protection of certain rights"[38] – related to the safeguarding of, amongst other, public order, child protection, and protection from defamation. Core to this Code is the possibility for the government to order the shutdown of any media publication that poses a risk to the protection of public order. Namely, any publication that undermines, amongst others, the Kingdom's territorial integrity, insult or offence the monarchy, or incite to discrimination or hatred.

Whilst the law is targeted at new outlets, some of its provisions also cover public speech. Similar to the Anti-Terrorism law, the Press Code sanctions the public diffusion, by any individual, of speech and content that disrupt "public order", including incitement to and apology of terrorism. Furthermore, the provisions related to publications shared by online means have been criticised on the basis that it could be interpreted in a broad manner, thus allowing for the sanctioning of online content in general. The 2016 Press Code, contrary to its predecessor, does not sanction any speech offence by jail time. However, it states that certain journalistic content, including that deemed to condone or incite to terrorism, can be ordered to be removed by judicial authorities.

The Moroccan Penal Code lists similar limits to freedom of speech than the Press Code. However, and in a major difference, speech offences under the penal code can be penalised with jail times. Specifically, anyone who undermines "the Islamic religion, the monarchy or incites to undermine the territorial integrity of the Kingdom."[39] can face up to 2 years imprisonment and a 200.000 Dirham fine (around 20,000 USD). This provision, under Article 267-5, also applies to speech and content shared online.

[38] "De la protection spéciale de certains droits"
[39] "quiconque porte atteinte à la région islamique, au régime monarchique ou incite à porter atteinte à l'intégrité territoriale du Royaume."

Civil society groups, including Human Rights Watch and Freedom House, have raised concerns with the Penal Code being used to silence non-violent speech that criticises the government and the monarchy online – including content shared on social media and video hosting platforms such as Youtube. A February 2020 report by Human Rights Watch (HRW) underlines rising concerns for online freedom of expression in Morocco. According to HRW, "at least 10 activists, artists, or other citizens who did nothing but peacefully express critical opinions via Facebook posts, YouTube videos, or rap songs" were arrested in the period between September 2019 and the publication of the report.

Draft social medial law: from consumer reviews to 24h removal deadline.

In March 2020, the Government Council approved draft law No 22.20, the so-called "social media law". A backlash on social media ensued with the draft being leaked online by Mustapha Swing, an online content creator. The leak, to this day the only "public" version of the draft law, revealed that online users calling for the boycott of certain products could face not only fines but also up three years of jail time, and that ISPs would be required to register to be allowed to operate in the Kingdom. Platforms failing to comply with the requirements laid out in the draft law could face administrative fines, a temporary suspension of their services, as well as the risk of their operating license being withdrawn.

Whilst the draft law mostly stems from a 2018 boycott of Afriquia Gaz, Centrale Danone and Sidi Alii products,[40] certain of its provisions could have significant impacts for online freedom of expression. Article19 and MENA Rights Group further shed light on this via a legal analysis shared with the UN Special Rapporteur on freedom of opinion and expression in June 2020. The groups focus on Articles 8, 10, 11 and 12 of the draft law. Article 8 would grant "broad censorship powers to network providers" by requesting ISPs to remove and restrict access to online content that "constitutes a dangerous threat to security, public order or which would be likely to undermine the constants of the Kingdom" within 24 hours. The language used here to sanction certain online content is further extending the provisions used in the Press and Penal Codes. A short removal deadline that, according to Article19, does not allow for proper assessment of the illegality of online content and that risks platforms over censoring content to comply with the law. Further, Article19 notes that delegating the adjudication of illegality of online content to online platforms is contrary to international standards and risk the "privatisation of judicial prerogatives". Article19 also recalls the French Constitutional Council's recent censuring of the "cyber-hate law" which included similar provisions. On Articles 10, 11, and 12, Article19 is particularly critical of the law establishing a "control body" without providing any further information regarding said body and its establishment.

Following public backlash on social media, the draft law was suspended in June 2020. With the suspension said to last until the end of the Covid-19 crisis, Article19 has called for the government to hold a public consultation, with different stakeholders, to "develop a legislative framework for the

[40] In April 2018, these three companies, some of the most important ones in the country and symbolising "an economy dominated by large groups linked to a business and political elite, or foreign brands", were targeted by an important boycott campaign calling for the high price of these products to be reduced as Moroccans were facing dire economic hardship. The campaign, called for by civil society organisations on social media, soon became the most popular boycott campaign in the country. Ultimately turning into a broader protest against socio-political injustice in Morocco.
See: "Let it Spoil!" Morocco's Boycott and the Empowerment of 'Regular' Citizen, AlJazeera Centre for Studies, November 2018

use of social networks that complies with international standards", as well as resolve the lack of transparency that has characterised this law until now.

The draft law is reminiscent of the attempt at creating a Numeric Code in 2013. This proposal to "structure" the online space included strict jail terms for a broad range of online content, including content contrary to public order, common decency or undermining Islamic religion. The draft code was met with public backlash before being overturned, and all copies of the code were entirely removed from the internet by the government.

Resources

Ait el Haj 2013), *Le code numérique avorté!,* L'Economiste.

Article19 (2020), *Morocco: government must fully withdraw draft law on social media.*

APA News (2020), *Maroc : Un projet de loi sur l'utilisation des réseaux sociaux suscite l'ire de l'opinion publique*

Bouhrara (2020), *Projet De Loi 22.20 : Le Gouvernement Tenterait-Il De Nous Faire Porter Des Muselières ?,* EcoActu.

L'Economiste (2020), *Loi 22.20: le projet de loi reporté jusqu'à la fin de l'urgence sanitaire.*

El Khamlichi (2020), *Projet de loi 22.20 : Quid des droits du consommateur marocain ?,* Maroc Diplomatique.

Freedom House (2020), *Freedom on the Net: Morocco.*

Human Rights Watch (2005), *Morocco's Truth Commission: Honoring Past Victims during an Uncertain Present.*

Human Rights Watch (2020), Morocco: Crackdown on Social Media Critics.

Iraqi (2018), *Ce jour-là : 16 mai 2003, les attentats de Casablanca*, Jeune Afrique.

Maghraoui (2008), *Morocco's Reforms after the Casablanca Bombings*, Carnegie Endowment.

Le Matin.MA (2020), *Le projet de loi sur l'utilisation des réseaux sociaux adopté en Conseil de gouvernement*

Oudrhiri (2020), *Projet de loi sur l'utilisation des réseaux sociaux, un nouveau boulet pour l'Exécutif ?,* TELQUEL.

Perspective Monde, *16 Mai 2003: Attentats terroristes à Casablanca, au Maroc*, Sherbrook University, Quebec.

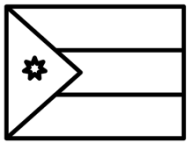
Tech Against Terrorism (2020), *The Online Regulation Series: France.*

Zaireg (2013), *Trois vérités toujours bonnes à rappeler sur le code numérique,* TELQUEL.

Royaume du Maroc — Ministère de la Justice et des Libertés, *Code Penal, version consolidé en date du 5 Juillet 2018.*

Royaume du Maroc – Bulletin Officiel (2016), *Dahir n° 1-16-122 du 6 kaada 1437 (10 août 2016) portant promulgation de la loi n° 88-13 relative à la presse et à l'édition*

MENA AND SUB-SAHARAN AFRICA | JORDAN



Jordan's online regulatory framework consists of four sets of legislation: anti-terrorism laws, cyber security regulation, cybercrime laws and the Telecommunications Act. Together, they regulate online content, and particularly terrorist use of the internet. On the whole, Jordan puts the emphasis on internet users, rather than on imposing requirements on the tech companies.

Jordan's regulatory framework:

- *The Anti-Terrorism Law No. 55 2006* (also called the Prevention of Terrorism Act), provides a definition of terrorism and criminalises related offences such as terrorist financing, terrorist recruitment, and establishing a group with the aim of committing terrorist acts.
- *The Anti-Terrorism Law 2014* amends and replaces four articles in the Anti-Terrorism Law 2006 to widen the definitional scope of terrorism to include any act that distorts the public order or harms Jordan's relationship with foreign countries. It also adds that an "information system or network" that supports, or spreads ideas of a terrorist group constitutes terrorism.
- *The Cybercrime Law 2019* criminalises hate speech as well as "fake news".
    - The law was based on the Cybercrime Law 2015, a draft law that was withdrawn from parliament in order to modify and align the law with existing penal codes.[41]
- *The Jordan Information Systems and Cyber Crime Law 2010* , also called the *Cyber security law*, is the first Jordanian law on cybercrimes and criminalises offences committed through the use of computer and electronic devices. Section 10 details the crime of the promotion and facilitation of terrorism through online means.
- *The Telecommunication Act 1995,* regulates all telecommunication companies in Jordan and establishes the regulator. The Act criminalises the "illegal" use of a public or private telecommunications network, as stipulated by Jordan penal codes.

Main regulatory body:

- The Telecommunications Regulatory Commission is responsible for regulating telecommunications and information technologies. They set the policies operators need to comply with and grant licenses. As part of their mandate within the Cybercrime Law 2019 they also oversee "applications". (I.e. apps).
- The Media Commission regulates broadcasting media and can shut down websites that have committed, or are suspected of committing, an offence as stipulated by Jordan penal codes.

[41] There has been a lot of criticism on this procedure, as several civil society groups such as AccessNow have argued that the law was withdrawn so that new amendments could be added before introducing it to parliament. AccessNow (2019). Cybercrime law in Jordan: Pushing Back on New Amendments that Could Harm Free Expression and Violate Privacy.

Key bodies and institutions:

- The Ministry of Information and Communications Technology (MoICT) sets the policy directions for telecommunications and information technologies through biannual national strategic plans, coordinates with relevant stakeholders, and submits policies to the Council of Ministers for approval.
- The Military State Security Court tries all individuals for terrorist offences, including terrorist use of the internet, as adopted in the Cybercrime Laws.
- The General Intelligence Department (GID) ("Da'irat al-Mukhabarat al-'Amma") – the country's intelligence agency which is in charge of intelligence operations to safeguard regarding national security. They are also involved with detaining individuals on the basis of the legal acts in question and the monitoring of suspects, both offline and online.

Main takeaways for tech companies:

The Cybercrime Laws

- *The Cybercrime Law 2019* puts liability on internet users for user-generated content, with the subsequent enforcement system serving users with prison time, or fines.
- Article 2 adds "applications" to the definition of telecommunications, therefore placing messaging apps in the remit of both the cybercrime laws and the Telecommunications Act, meaning that the law now applies to smart phone apps.
    - This means that they now also need to comply with Article 29 of the Telecommunications Act, by allowing the monitoring of telecommunication entities when suspecting of committing a crime (see below)
- The Cybercrime Law 2019 considers any media or publishing material that "facilitates the commission and promotion of terrorist acts" to be terrorism. This can include any website or media company that enters into such action

The 1995 Telecommunications Act

- The *1995 Telecommunications Act* defines a telecommunications service and, in article 29, stipulates that the telecommunications service needs to allow relevant authorities to monitor their users' communications. Therefore, all providers can be asked to share information on their users with legitimate authorities (such as the GID).
- When a website (whether a service provider, operator, or application) commits or is suspected of committing an offence under the Jordan penal codes, the Media Commission or the government can shut down a website or interrupt its services

Counterterrorism legislation

The Anti-Terrorism Law of 2006 was the first piece of counterterrorism legislation introduced by Jordan, following the 2005 hotel bombings in Amman, Jordan, committed by al-Qaeda in Iraq. The law is the cornerstone of terrorism legislation in Jordan, and puts any offences related to terrorism in the jurisdiction of the Military State Court. The law criminalises the support of a terrorist group, the funding of a terrorist group, and the creation of a terrorist group. The amendment also widens the definitional scope of terrorism in Jordan's penal code, to include any media or publishing material that "facilitates the commission and promotion of terrorist acts". Furthermore, the law allows for the general prosecutor to monitor any individual that is suspected of terrorist offences. Finally, the law also imposes criminal penalties where the police can detain anyone suspecting of spreading hate speech for 24 hours to 7 days, subject to extension for a period of one month.

The widening the definitional scope of terrorism was met with criticism from civil society groups, such as Human Rights Watch, pointing out how this definition can be used to quell not just expressions of terrorism, but also peaceful and legal speech with no relation to terrorist organisations. In addition, Open Democracy criticised the law for having been used in prosecutions of human rights activists and journalists. The United Nations Educational, Scientific and Cultural Organisation pointed out that prosecuting media workers in military courts, and allowing detention pending trial, is particularly problematic for freedom of expression, as it incentivises people to refrain from discussing anything that might be considered terrorist in nature.

Cybercrime legislation

The Cybercrime Law of 2019 criminalises hate speech, defined as “every writing and every speech or action intended to provoke sectarian or racial sedition, advocate violence or foster conflict between followers of different religions and various components of the nation.”[42] Those found guilty risk facing a 3 month to 3-year prison sentence, as well as a fine of 1,000 to 3,000 Dinars (1410 - 4231.36 USD) The law was met with criticism by the public, sparking a media campaign on Facebook and Twitter, called "#withdraw_cybercrime_law" , where concerns of freedom of speech were the main drive behind the movement.

Civil society groups such as AccessNow mirror the initial concern over the law and argue that the definition of hate speech is too broad, and likely to apply to online content that might not incite hatred or harm. AccessNow deems the law to smudge the line between hate speech and what can be considered legal criticism of Jordanian officials online and argues this might lead to the censorship of activists. This particular point was highlighted by the United States Justice Department under its Freedom of the Net branch, which showed how in 2019 numerous activists were arrested and prosecuted for their social media posts.

The enforcement mechanism behind the law has also received criticism. Namely, the details of anyone suspected of terrorist activity online, can be requested from an “application” or an Internet café. In terms of the former, article 2 of the 2019 amendment stipulates that “applications” fall under the definition of an information system, which according to article 29 of the Telecommunications Act, can be monitored when legitimate authorities deem this to be appropriate, without a court order. Human Rights Watch points out that this might lead to individuals being prosecuted for their private conversations. Access Now, on their part argued that the law can be used for “mass surveillance” through monitoring messaging apps. In terms of the latter, all Internet cafes need to keep tabs on who uses their Internet services and are required to keep the records for 6 months.

Future Developments of Jordan’s regulatory framework

Jordan periodically reviews its cyber legislation, as shown by the many amendments throughout the years. Barring criticism regarding freedom of speech, the Middle East Institute, a non-profit thinktank, recognises the Jordan government’s policymaking as adaptive and has praised Jordan’s review process to try and ensure its cybersecurity legislation stays up to date in an evolving cyber security landscape.

Finally, The Middle East Institute reports that the 2019 cybercrime law announced two additional new structures, the National Cybersecurity Council and the National Center for Cybersecurity, however at this time of writing no further information on these entities could be found.

[42] As translated by Access Now, to be found here Cybercrime law in Jordan: Pushing Back on New Amendments that Could Harm Free Expression and Violate Privacy.

Resources

AccessNow (2019), *Cybercrime law in Jordan: Pushing Back on New Amendments that Could Harm Free Expression and Violate Privacy.*

Alkarama Foundation (2017), *Jordan Shadow Report - Submitted to the Human Rights Committee in the Context of the Review of the Fifth Periodic Report of Jordan*.

Center for Defending Freedom of Journalists (2014) CDFJ Launches its 2014 Annual Report on Media Freedom Status in Jordan - "Dead End".

Freedom of the Net (2019) *Jordan.*

Human Rights Watch (2014), *Jordan: Terrorism Amendments Threaten Rights.*

Human Rights Watch (2019) Jordan: 'Fake News' Amendments Need Revision.

Osman (2016), *10 Years On: Jordan's Anti-Terrorism Law and the Crackdown on Dissent*, Open Democracy.

King Hussein Foundation. – Information and Research Center (2015), *A Glimpse into the Perception of Digital Privacy in Jordan*.

Ghazal (2018), *New Cybercrime Law will Restrict Media Freedom, Public Opinion*, The Jordan Times.

Omari (2018), *Jordanians launch social media campaign against new cybercrime law,* The Jordan Times.

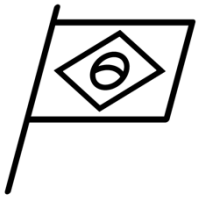
Faqir (2013), *Cyber Crimes in Jordan: A legal Assessment on the Effectiveness of Information System Crimes Law No (30) of 2010*, International Journal of Cyber Criminology.

Araz (2020), *Jordan Adopts Sweeping Cybersecurity Legislation,* The Middle East Institute.

United Nations Educational, Scientific and Cultural Organization. – The International Programme for the Development of Communication, (2015) *Assessment of Media Development in Jordan.*

Telecommunications Regulatory Commission Jordan, *Telecommunications Law No. (13) of 1995.*

LATN AMERICA | BRAZIL



Brazil represents a major market for online platforms. It is the leading country in terms of internet use in South America, and a key market for Facebook and WhatsApp. WhatsApp's popularity and the online disinformation campaigns that have been coordinated on the platform are essential to understanding Brazil's approach to online regulation. The messaging app has been accused of being used "for the dissemination of fake news", whilst critics of the country's "fake news" bill have said that it served as a "standard" for new regulation in the country based on the app's existing limitations on forwarding messages and group size.

Brazil's regulatory framework:

- *Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet*, the Brazilian Internet Freedom, Responsibility and Transparency Act, or Law PLS2630/2020, passed by the Senate earlier this year, and yet to be approved by the Chambre of Deputies before being signed into law by President Bolsonaro.
    - The law is meant to counter the spread of misinformation online and would oblige messenger apps to implement measures to ensure the traceability of messages shared, as well as compel tech platforms to monitor inauthentic behaviour.
- *Marco Civil da Internet* (MCI), also known as the Brazilian Internet Bill of Rights, passed in 2014 and fully implemented in 2016, "modifies the country's constitution to give citizens, the government and organizations rights and responsibilities with regard to the Internet".
    - The bill lays out a set of 10 principles for the governance of online networks in Brazil, including: "network neutrality, privacy, freedom of expression, security and universality."[43]
    - The bill underwent a long process of discussions and reviews, involving individuals, organisations, tech platforms, and other governments between 2009 and 2014.
    - The MCI makes Brazil one the largest countries in the world where the "democratic norm of equal access to information online" is inscribed in its Civil Code.
- *Lei nº 13.260, de 16 de Março de 2016,* Brazil's Anti-Terrorism law, amended in 2016 ahead of the Olympic Games.
    - The law does not address use of the internet for terrorist purposes, but covers issues related to the promotion and preparation of terrorism.

[43] The principles of universality mainly relates to ensuring the diversity of internet users and "spurring innovation"

Key takeaways for tech platforms:

- The MCI exempts platforms from liability for user-generated content, unless in cases where a court order states the content was illegal, in which case platforms must remove the content or face legal liability.
- Defamation has been used as a basis for judicial authorities ordering removals of online content.
- Violations of electoral laws have also been used as a motivation for removal orders.
- Brazil's "fake news" bill, once passed into a law by Presidential decree, would have major consequences for online platforms, especially encrypted messaging services, by imposing:
    - Traceability requirements for messaging services: messaging apps would be required to store the logs of "broadcasted messages", meaning messages sent by more than 5 users and reaching a least a 1,000, for three months .This requirement is linked to a "technical capability directive" for platforms to be able to trace back individual messages.
    - Reports of "automated or inauthentic accounts", or the reception of a court order, would require online platforms to confirm the identify of their users.

A misinformation problem

Brazil has, in recent years, seen significant spread of dis- and misinformation on social media, in particular during the 2018 Presidential elections and the Covid-19 crisis. The debate around fake news and online platforms in Brazil has particularly focussed on WhatsApp, the most popular messaging app in the country, on "malicious coordinated action" to spread misinformation on the app, and on how to counter this via traceability requirements.

Judicial authorities in Brazil have the power to order the removal of certain content and blocking of accounts on the basis of defamation or violations of election laws. This is one of the most important exceptions to the tech companies' protection from legal liability for user-generated content granted by the MCI. Just this August, Facebook was forced to comply with a Supreme Court removal order, for accounts spreading false information about Brazilian judges. The company did so in part to prevent one of its employees from facing potential criminal liability. The company was nonetheless fined around $368,000 for not blocking the accounts worldwide.[44]

Brazilian Internet Freedom, Responsibility and Transparency Act, towards a new framework to tackle fake news

Beyond judicial orders for account and content removal, Brazil vowed to tackle its misinformation problem with a so-called "fake news" bill: Law PLS2630/2020, or the Brazilian Internet Freedom, Responsibility and Transparency Act. If signed, the law could bypass some of the principles set out in the MCI, in particular by making companies legally liable for content published on their platforms,

[44] Twitter was also ordered to take down accounts, and complied with the order whilst saying it would be appealing it.

and thus acting as “a powerful incentive to limit Brazilians’ freedom of speech [at a time of political unrest].”

The initial proposal included substantial provisions on data retention for messages that would have met a “virality threshold” of being forwarded by more than 5 users to 1,000 users within 15 days, as well as a requirement for tech companies to verify the identity of their users. These original requirements were amended in the version approved by the Senate, however, the bill will still impose important traceability and monitoring obligations on tech platforms in order to detect bot accounts and inauthentic behaviours that spread fake news.

The requirement to retain chains of communications, for instance, is still present in the current version of the bill, though limited to “private messaging applications” and only requiring companies to store the logs for three months. This requirement has been criticised by the Electronic Frontier Foundation (EFF) for imposing a “tech mandate” on messenger apps that would weaken privacy protections by compelling them to retain important chain of communications. Attached to that is the requirement for messaging apps to limit the size of their private groups and lists. A limitation that WhatsApp already implements and that has led some commentators to criticise the bill for creating a “standard based on WhatsApp” for its group size limits and limitation on forwarding messages.

Although the current version of the law has removed the requirement for “large” social media and messaging apps to collect users legal identification information (national identity cards) to use their services, platforms “may” still be required to confirm users identify following “reports of non-compliance with the fake news law, evidence of automated or inauthentic accounts, or upon court order”. With the threshold for what constitutes such reports being unclear, civil society organisations, including the Center for Democracy & Technology (CDT), have expressed concerns that this would lead to “arbitrary” and “excessive” violation of users’ privacy and right to freedom of expression. Tech companies will also be required to ensure that their services are not used for inauthentic behaviour and develop the necessary “technical means” to do so. The monitoring of such “inauthentic behaviour” and the technical means it will require have been criticised by the CDT on the grounds that the law “is unclear and poses a potential threat to online privacy and security.”

The EFF has also criticised Article 37 of the law, which requires platforms to appoint a representative in Brazil and ensure that users databases can be accessed by the staff in the country, in case they would be required to hand them over to law enforcement. Another concern raised by the EFF regards the wide application of the law beyond Brazil. Indeed, the EFF has criticised it for not being limited to online services in Brazil, but that it is said to apply at “company level” without regard for where the user is located, or their nationality.

With the laws having been criticised for the unproportionate risks it poses to freedom of expression and users right to privacy, as well as for impending innovation by compelling platforms to change how they store messages, Brazilian digital rights experts have also denounced it for failing to meet its said aim of tackling misinformation. Rolando Lemos – a Brazilian lawyer, digital rights expert and member of the Facebook Oversight Board – has criticised it for focusing too much on content and failing to counter the professional networks of disinformation: “it attacks the leaves and not the root

of the problem, which is the fight against those who finance disinformation campaigns in a hidden way".

Resources

AccessNow (2020), *Brazil Congress moving forward disinformation bill that brings free expression and privacy harms to new levels.*

Al-Jazeera (2020), *Facebook bows to Brazil court order, bans pro-Bolsonaro profiles.*

Arnaudo (2017), Brazil, the Internet and the Digital Bill of Rights: Reviewing the State of Brazilian Internet Governance, Igarape Institute

Counter Extremism Project, *Brazil: Extremism & Counter-Extremism.*

Garcia Tsavkko (2020), *Brazil's "fake news" bill won't solve its misinformation problem.* MIT Technology Review.

Isaac and Roose (2018), *Disinformation Spreads on WhatsApp Ahead of Brazilian Election,* The New York Times.

Human Rights Watch (2015), *Brazil as the Global Guardian of Internet Freedom?*

Fleischmann Do Amaral (2020), *What is the proposed fake news regulation in Brazil and how does it affect social media in the country?,* LABS

Freedom House (2020), *Freedom on the Net: Brazil.*

Lima (2020), *Misinformation campaigns escalate during Covid-19 pandemic in Brazil.*

Lyons (2020), *Brazil Supreme Court orders Facebook to block accounts of several Bolsonaro allies,* The Verge.

Maheswar and Nojeim (2020), *Update on Brazil's Fake News Bill: The Draft Approved by the Senate Continues to Jeopardize Users' Rights*, Center for Democracy and Technology.

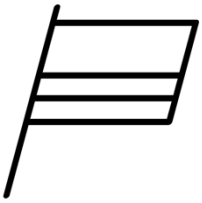
Rodriguez and Schoen (2020), *FAQ: Why Brazil's Plan to Mandate Traceability in Private Messaging Apps Will Break User's Expectation of Privacy and Security*, Electronic Frontier Foundation.

Rodriguez and Schoen (2020), 5 *Serious Flaws in the New Brazilian "Fake News" Bill that Will Undermine Human Rights,* Electronic Frontier Foundation.

Saraiva (2020), *Tackling Disinformation in Brazil,* Foreign Policy.

Uchoa (2020), *Brazil coronavirus: 'Our biggest problem is fake news'*, BBC News.

LATIN AMERICA | COLOMBIA



With a growing internet penetration rate (69%) and an increasing number of active social media users (35 million, at a growth rate of 11% between 2019 and 2020), the online space in Colombia remains governed by the principle of net neutrality.

This principle is enshrined in the country's legal framework, in particular in Article 56 of Law 1450 of 2011 which serves as a framework for the guarantees and responsibilities of the states towards its citizens. In effect, the principle of net neutrality in Colombia serves as the basis for justifying the non-discrimination of online content and services, and has been invoked by the Ministry of Information and Communication to justify the non-blocking of apps in the country. As a result, only child sexual abuse material is considered illegal online content under Colombian law, and it is systematically blocked in the country.
However, a decision made in December 2019 by the Colombian Supreme Court could significantly change the country's online landscape. Ruling on the protection of a person's reputation online, the Supreme Court stated that blog operators could face legal liability if they failed to adopt proper moderation mechanisms for the comments published on their sites and online forums. These mechanisms should also include systems to identify the author of a post, thus lifting the possibility of online anonymity.

This decision by the supreme court has been criticised by civil society organisations, including the Fundación Para la Libertad de Prensa (FLIP, Foundation for the freedom of the press), which in its 2019 report on the state of the internet, *El Internet que Nadie Querie,* underlined that this decision was one amongst other legislative proposals that were leaning towards a more restrictive online space, and presented risks for online freedom of expression. With regard to the Supreme Court Decision of December 2019, the FLIP noted that: "The decision is dangerous for freedom of expression since, by holding the media or blog operators responsible for what is published by their users, an incentive is created for those to excessively restrict comments or completely eliminate these sections for fear of eventual legal consequences."[45]

In this same report, the FLIP shows a trend for more stringent online regulation in Colombia demonstrated by different legislative proposals, made between 2012 and 2019, that would have substantially limited freedom expression on the internet – and in some occasions failed to meet the Constitutional standards requirements in place in Colombia. Amongst the proposals underlined by FLIP, is bill 176/19 – presented in 2019 – which aimed at regulating the use of social media platforms. The proposed bill planned on doing so by requesting written consent to publish any type of information or data about a person (including photograph or video). The proposal also included provisions on the prohibition of insults, and on preventing people from "overexposing" their own privacy, or from accessing "inappropriate content" online – without defining such content.

[45] "La decisión es peligrosa para la libertad de expresión ya que, al hacer a los medios u operadores de blogs responsables de lo publicado por sus usuarios, se crea un incentivo para que aquellos restrinjan en exceso los comentarios o eliminen completamente estas secciones por temor a eventuales consecuencias legales."

Resources

Freedom House (2020), *Freedom on the Net: Colombia.*

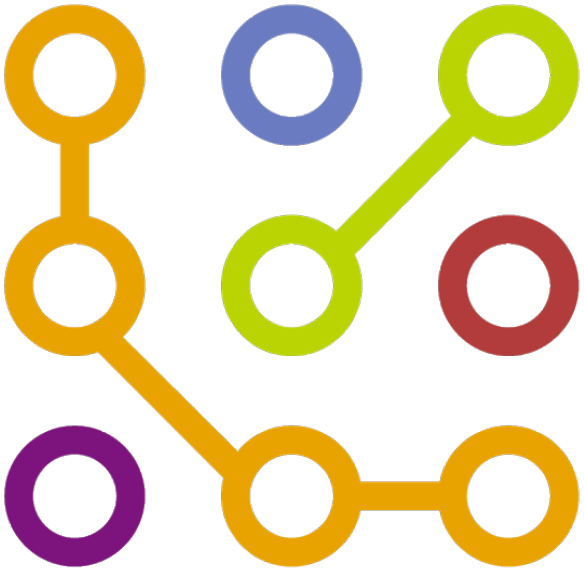
Fundación Para la Libertad de Prensa (2019), *El internet que nadie quiere,*

Henshaw (2020), *Online extremism in Latin America – An Overview,* GNET.

Henshaws (2020), *Extremist responses to covid-19 in Latin America,* GNET.

Legis Ambito Juridico (2019), *Operadores de blogs pueden ser civilmente responsables por contenidos difamatorios*

Wired (2020), *What is Net Neutrality? The Complete WIRED guide.*



Tech Against Terrorism connects industry, government, and civil society to prevent the terrorist use of the internet whilst respecting human rights

A project supported by UN CTED
under mandate of the United Nations Security Council Counter-Terrorism Committee

techagainst**terrorism**.org @techvsterrorism
contact@techagainstterrorism.org