tech
against
terrorism

THE ONLINE REGULATION SERIES
| THE HANDBOOK

June 2021

CONTENTS

INTRODUCTION

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.

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.

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.

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.

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. In July 2021, we launched an updated version of the Knowledge Sharing Platform, which collates tools and resources to support tech companies in tackling terrorist use of the internet.

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.
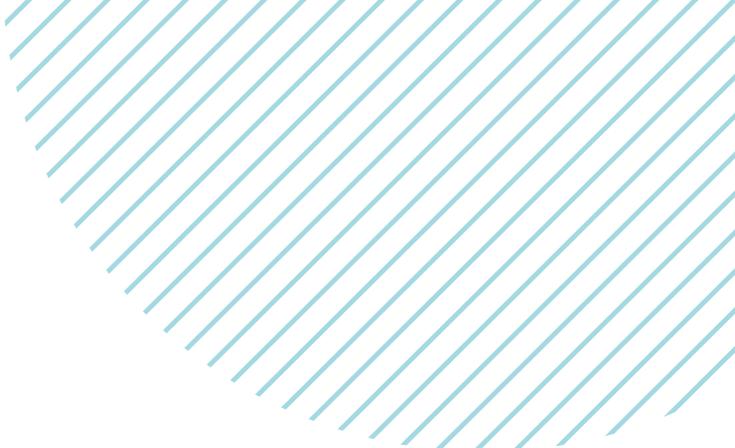
Our mission is to support smaller tech companies in tackling this threat whilst respecting human rights, and to provide companies with practical tools to facilitate this process.

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 and our current work on the Terrorist Content Analytics Platform.

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.

BACKGROUND TO THE ONLINE REGULATION SERIES

Since 2017 there have been many developments in the regulation of online speech and content, in particular in how we counter the spread of terrorist content online. Several new laws have been passed or proposed in countries like 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. We reviewed over 60 pieces of legislation, proposals, and guidelines that aim to regulate the online sphere, and analysed over 100 data sources and civil society reports.

This effort culminated in the Online Regulation Series, where over the course of six weeks, in October and November 2020, Tech Against Terrorism focused its outreach and knowledge-sharing efforts on providing our stakeholders with an update on the state of global online regulation.

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 aim to regulate online content?
- What are the implications for tech platforms?

Throughout the series, we published 20 blogposts on our website, sharing relevant resources and insights on Twitter as well. The series covered:

- 17 country-specific 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, bringing together analysis from tech policy and digital rights experts on the key global regulations that are shaping online speech around the world.

Editorial note: The analysis included in this report is based on the blogposts we published on Tech Against Terrorism's website in October – November 2020, and were updated to reflect changes in the online regulation landscape that took place between October 2020 and June 2021. 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.

If you are aware of something that should be included or updated, please get in touch with us at contact@techagainstterrorism.org

THE ONLINE REGULATION SERIES | OVERVIEW

When conducting our research for the Online Regulation Series, we identified three separate regulatory aims used by governments to justify regulating online content:
Countering terrorist and violent extremist content, or "harmful" content:

These regulations target terrorist use of the internet by compelling tech companies to rapidly remove terrorist and violent extremist content from their platforms, often including short removal deadlines (from 1 to 24 hours) and heavy fines in cases of non-compliance. The German NetzDG (2017) was the first of such regulations and was followed by similar moves in other jurisdictions, including France, the UK, the EU, and more recently Canada. Some of these laws also target "harmful" online content more generally, which can span anything from illegal content and incitement to hatred to suicide-promoting content.

Countering the spread of Misinformation and Disinformation:

In some countries, policymakers have focused regulatory proposals on misinformation and disinformation. These proposals often include the power for governments to issue removal or correction orders to platforms, as is the case in Singapore, or the power for platforms to trace the originator of a message, as has been introduced in India and discussed in Brazil.

Adapting to the Digital Space:

These laws 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 EU's Digital Services Act has been explicitly framed as a response to how digital changes impact our lives. Canada's Communications Future: Time to Act report also outlined recommendations for a thorough change to the country's regulation of online platforms and content.

THE ONLINE REGULATION SERIES |
MAP OF REGULATORY THEMES
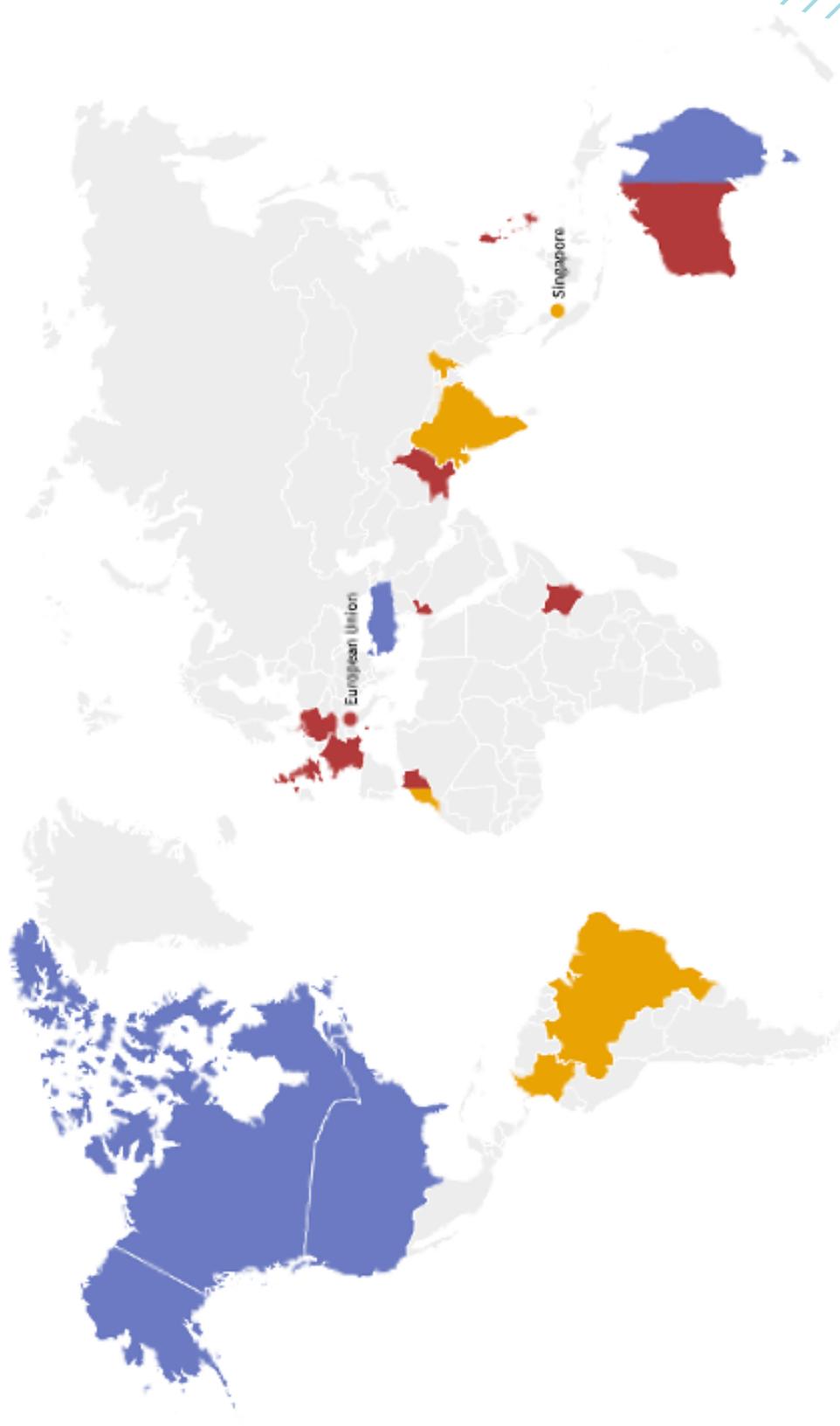
SECTION 1 |
THE STATE OF ONLINE REGULATION

TECH AGAINST TERRORISM | RECOMMENDATIONS FOR GOVERNMENTS

Based on our analysis of existing and upcoming regulations aimed at countering terrorist and other harmful content online, we call on governments to:

1. Safeguard the rule of law

- Avoid measures that risk undermining the rule of law and due process. In particular governments should:
- Ensure that definitions of key terms, such as terrorist content, are clear, practical, and have a basis in existing legal frameworks. Governments should also avoid introducing regulation that depends on subjective interpretation of harm, as this is often difficult for tech companies to operationalise at scale without negatively impacting freedom of expression.
- Use legal powers to promote the rule of law through more comprehensive terrorism designation lists (in particular of far-right terrorist groups) to help increase definitional clarity around terms such as terrorism. Refrain from making content that is legal online, illegal offline. There should be a clear legal basis to remove online content, including via existing counterterrorism laws and terrorism designation lists, or via existing limitations to freedom of expression.
- Refrain from introducing provisions that infringe on existing due process with regards to limitations to freedom of expression. In line with international human rights standards, limits to freedom of expression should be adjudicated by an independent judiciary body and not delegated to a private entity.
- Provide legal certainty to tech platforms by clarifying how regulatory compliance will be assessed, and by providing guidance on the specific steps companies should take to comply with legal requirements

2. Consider the capacity and resources of smaller platforms, respecting the principle of proportional regulations and the principle of equality before public charges.

- Ensure obligations for tech companies are proportionate according to size and capacity, and avoid harming competition and innovation by limiting financial penalties for smaller or micro-platforms.
- Increase support for the tech sector, particularly for smaller platforms, in countering terrorist and violent extremist use of the internet, for example through public-private partnership endeavours, and digital literacy programmes. We know from experience that smaller platforms are very receptive to mentoring and any opportunity to learn how to minimise the terrorist and violent extremist threat online. If governments wish to tackle online harms – including terrorist content – effectively, we recommend they invest in similar programmes to support smaller platforms.

3. Provide clarity regarding the safeguards and redress mechanisms

- Clarify what safeguards are in place to avoid removal of legal content.
- Clarify what redress mechanisms are in place in case of erroneous removal, in particular regarding content removal following removal requests from a country's judicial or governmental authority.

To ensure that human rights, in particular freedom of expression, are safeguarded when implementing online regulations, we call on governments to:

- Provide information on the steps taken by the relevant implementing and supervising authority to ensure that their mandates are carried out with the fullest respect for freedom of expression and human rights, and that they are:

  - fully aware of risks to human rights and freedom of expression associated with the measures they implement, for example removal orders and requirement to remove content within a specified timeframe.
  - uniform in their judgement and do not politicise removal orders.
  - consistent and accurate in issuing penalties to companies.
  - disincentivised from over-zealous content removal.
  - held accountable for assessments and judgements made in implementing this regulation.

4. Produce transparency reports on their engagement with tech companies for counterterrorism purposes, in line with the Tech Against Terrorism Guidelines[1]

Some of the online regulations that have been passed, or are being discussed at the time of writing, include provisions that Tech Against Terrorism strongly advises against. For governments that decide to pursue these provisions, we recommend the following:

Tech Against Terrorism argues that adjudication of the legality or harmfulness of content should be the role of governments, not tech platforms. For regulations that places the onus of adjudication on tech companies, we recommend governments to:

- Avoid introducing measures that do not allow sufficient time for platforms to adequately assess the legality of content, and provide the necessary practical support for platforms to correctly assess content.

Tech Against Terrorism strongly advises against placing liability for user-generated content on tech companies or their employees. If governments decide to pursue these liability regimes, we urge them to:

- Clarify under what exact circumstances a company's legal representative may be held liable for their company's lack of compliance with the regulation.

Tech Against Terrorism advises against mandating short removal deadlines for terrorist or harmful content, as these deadlines lack consideration for platforms' capacities and encourage overzealous removal of content. For governments that decide to mandate short removal deadlines, we call on them to:

- Consider the increase in resources (financial, human, and technical) these provisions require and small platforms' capacity.

---

1. Forthcoming 2021

We call on governments to take a holistic approach to countering terrorism and violence extremism. Beyond regulating terrorist and harmful content, governments should ensure that regulatory frameworks address the root causes of radicalisation and hold individuals that engage in terrorist and violent extremism activities accountable, in full respect for international human rights standards.
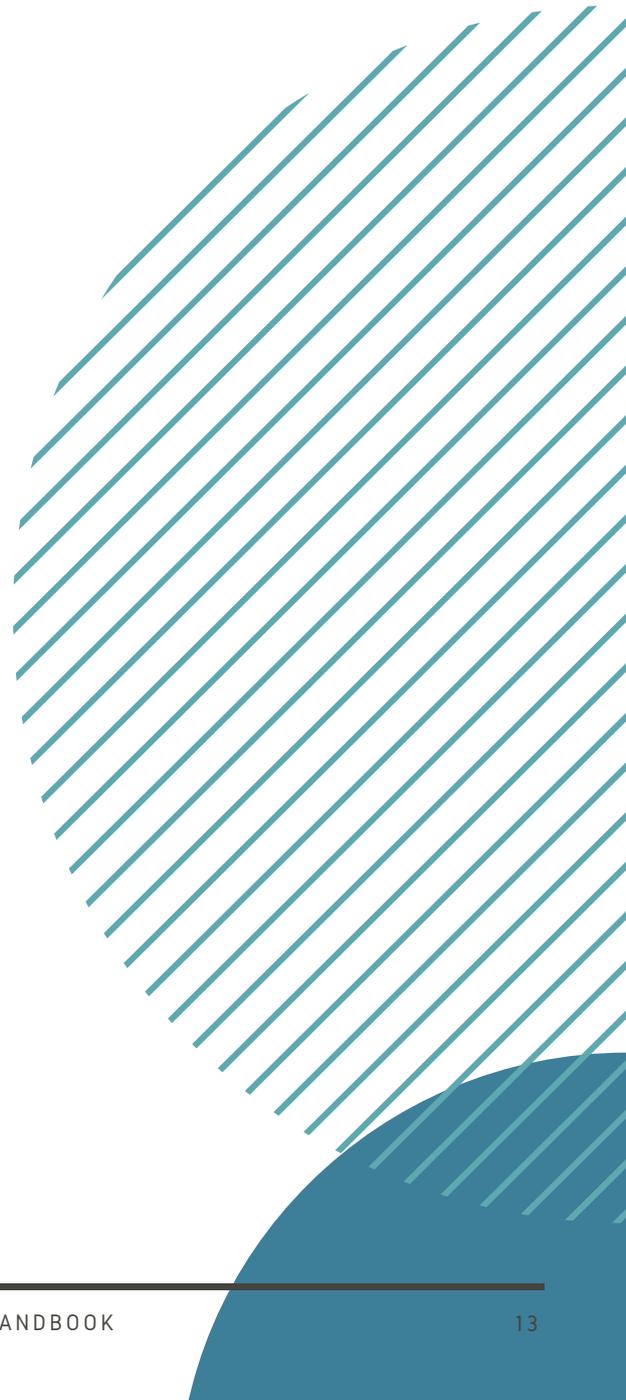
THE STATE OF ONLINE REGULATION | TECH AGAINST TERRORISM'S CONCERNS

Based on our analysis of online regulation globally and the regulatory key trends we identified, we develop in this section on our main concerns with the new wave of online regulation.

1. Lack of consideration for smaller platforms

Research conducted by Tech Against Terrorism has shown that smaller and newer tech companies are the most at risk of exploitation by terrorists and violent extremists. Most of the smaller platforms Tech Against Terrorism regularly engages with show willingness in tackling this threat but lack the human, technical, and financial resources required.

Despite this observation, most of the online regulations covered in this handbook apply indiscriminately to platforms of all sizes and resources. This means that small and micro-sized platforms are expected to comply with the same stringent legal requirements as larger and long-established platforms would do.

Such unrealistic expectations of compliance risk penalising small platforms with heavy fines and leaving them behind, instead of offering them the support needed to counter the threat. This also bears the risk of reduced competition in the tech sector if smaller platforms are not able to catch up or are financially compromised by the fines.

Based on our analysis of the regulations covered in this Handbook, we assess the laws passed in the following countries particularly lack consideration for smaller platforms:
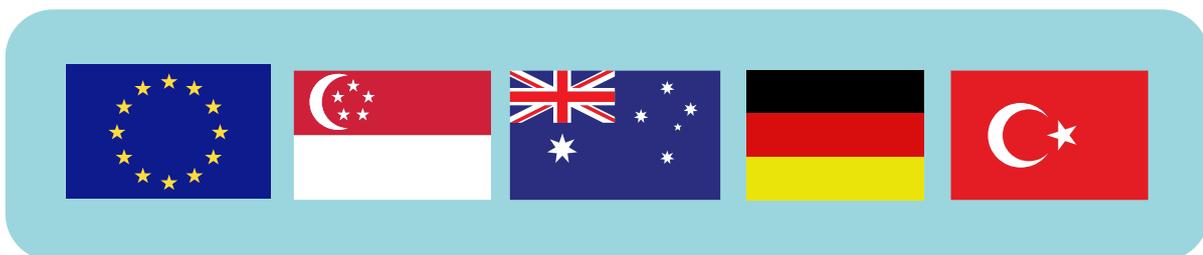
Recognising that not all platforms are equal in their capacity to comply:

Concerns regarding disparities in resources and how these impact a platform's capacity to comply with legal requirements were also raised by the French Constitutional Council in its censuring of the so-called "CyberHate" law. The Council stressed that some of the provisions in the original version law were impossible to satisfy and broke the principle of equality before public charges – which underlines that legal and administrative requirements should not cause heavy or particular burdens for those having to comply.

With this ruling, the Council recognised that platforms' resources can significantly impede their capacity to comply with legal requirements, and that requirements with high resource demands should not be included in online regulation.

Tech Against Terrorism urges policymakers to consider the diversity of platforms to ensure that the most demanding legal requirements consider platforms' sizes. In line with this, smaller tech companies should be consulted when new regulations are being drafted and discussed.

Policymakers should also support capacity-building and knowledge-sharing activities to strengthen smaller platforms' capacity to respond to terrorist and violent extremist use of the internet, and to comply with legal requirements.

Tech Against Terrorism works to ensure that smaller platforms are considered and heard. We regularly raise the importance of acknowledging that smaller platforms need additional support, rather than heavy fines, in our policy responses. To do so, we regularly consult with smaller tech companies engaged in our Mentorship and Membership programmes.

3. Lack of definitional clarity and risks for freedom of expression

Many of the regulations analysed for the 2020 Online Regulations Series are impractically broad in their definition of harmful content and circular in their explanation of terrorist content – they rarely indicate how to implement the definition of terrorism or harmful content. This presents serious risks for freedom of expression, as these regulations could be used to pressure tech companies to remove legal or non-violent speech.

With such vague definitions of "legal but harmful" content, countries are introducing mechanisms that risk undermining the rule of law. In a democracy, we cannot make speech that is legal offline illegal in the online space, and private organisations should not be pressured to remove legal content.

4. Online regulation and the risks of "censorship creep"

Danielle Citron (Professor at the University of Virginia School of Law and expert on information privacy and free expression), in her criticisms of the EU regulation of online content and EU Internet Referral Units, has expressed concerns with the risks of "Censorship Creep": "whereby a wide array of protected speech, including political criticism and newsworthy content, may end up being removed from online platforms on a global scale."

Citron's criticisms focus on "definitional ambiguity" around what constitutes harmful content, namely "hateful conduct" and "violent extremism material", which can be abused to target legitimate speech and political dissent. Combined with pressure on platforms to (rapidly) remove harmful content, this risks the over-removal of content which could have major repercussions on freedom of expression online.

Tech Against Terrorism cautions against vague and circular definitions of terrorist or harmful content in laws, and against governments demanding platforms to remove content that is not clearly prohibited by law. We call on governments to apply the same level of detail and clarity in their legislation that is expected of tech companies: clearly delineated and defined prohibitions, that are inscribed in the rule of law by reflecting behaviours and content that is illegal offline, instead of creating a differentiated regime for the online space.

TABLE |

Overview of countries aligning with the key trends identified by Tech Against Terrorism

In this table, "N/A" refers to the absence of a passed legislation aimed at regulating terrorist or harmful content online. These countries are considered in this Handbook due to regulatory discussion and legislative proposals, however, in the absence of a published draft bill, we refrained from classifying them in the below table.

Singapore and Jordan stand out in this table by being the only countries that do not follow any of the key trends Tech Against Terrorism identified. Please see our commentaries of each country to learn more about our analysis and assessments of online regulations in Singapore and Jordan.

KEY TRENDS IN ONLINE REGULATION

✓ Fully in line with the key trends identified
✓ Partially in line with the key trends identified (gray)

LEGISLATURES	Different requirements depending on platform size	Short removal deadlines	Increased reliance on automated moderation	Outsourcing legal adjudication	Platform employees liability or demanding a focal point	Mandating a local presence	Mandating transparency and accountability
Australia		✓		✓			✓
Brazil				✓ (partial)		✓	
Canada	Not applicable.						
Colombia	Not applicable.						
EU	✓ (partial)	✓	✓ (partial)	✓	✓	✓	✓
France	✓ (partial)					✓	✓
Germany		✓	✓ (partial)	✓			
India	✓ (partial)	✓	✓		✓	✓	✓
Jordan	The Cybercrime Law (2019) also applies online messaging services.						
Kenya						✓	
Morocco	Not applicable.						
Pakistan	✓ (partial)	✓	✓		✓	✓	
Philippines	Not applicable.						
Singapore	The Protection of Online Falsehoods and Manipulation Bill (2019), applies to all type of online platforms including encrypted messaging services.						
Turkey	✓	✓				✓	
UK				✓		✓	✓
USA	Not applicable.						

1. Short removal deadlines

Requiring smaller tech companies to remove content within short timeframes is a common f unrealistic expectation being placed on smaller companies in various jurisdictions. A one-hour deadline, for example, would likely require constant monitoring from tech platforms to ensure compliance. It is a difficult endeavour for most medium and large tech platforms and virtually impossible for smaller platforms.

The pressure put on tech companies to quickly respond to alerted content, and to proactively remove or prevent upload of content risks freedom of expression, as tech platforms will not have the time necessary to properly adjudicate on content legality. Instead, there is the risk of an overzealous removal of content, with platforms indiscriminately taking down all content reported for illegality or violation of the content standards, before properly reviewing a report.

The EU Regulation 2021/784 on Addressing the dissemination of terrorist content online, which mandates a one-hour removal for terrorist content for all platforms, has been amended in its final version to acknowledge that not all platforms have the same resources and capacities. Tech companies that cannot comply with a removal order will have to inform the competent authority of this without "undue delay", and will be excused if they can provide "objectively justifiable technical or operational reasons" as to why they cannot comply. However, this amendment still requires smaller tech platforms to rapidly acknowledge terrorist content alerts to avoid penalties, which does not resolve the issue of platforms having to be almost constantly monitoring alerts received.
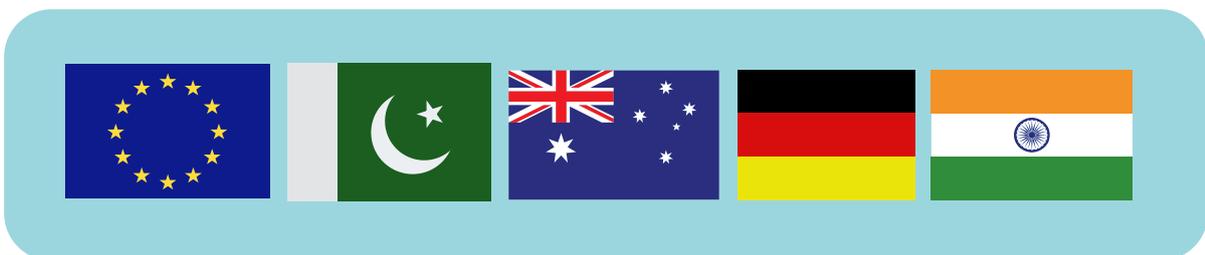
2. Increased reliance on automated moderation

For most platforms, stringent online regulation mandating content to be removed will require a significant increase in resources dedicated to content moderation. For the platforms that have technical resources, this will most likely mean increased reliance on automated content moderation tools.

Whilst automated content moderation has its benefits, current solutions are not nuanced enough to correctly assess whether certain pieces of content are in fact terrorist material or harmful. Most automated solutions notably lack the capacity to comprehend context (for example, whether content is journalistic or shared in order to criticise a specific position) and require human overview to avoid the excessive takedown of content. An increased reliance on automated moderation solutions raises the risk of false positives in taking down content that is legal, and raises questions about accountability in removal decisions. Our greatest concern is the risk that content denouncing human rights violations, including journalistic content that can serve as evidence of such violation, could be automatically removed, more so at a time where constitutional guarantees are weakened in certain countries.

Covid-19 and increased reliance on automated tools

YouTube's increased reliance on automated tools in 2020 demonstrates the risks of over-removing non-violating content. Due to the Covid-19 pandemic and ensuing lockdown measures, YouTube and many other large tech companies increased its use of automated moderation tools considerably. This resulted in more non-violating content being actioned, with the number of user appeals doubling and the number of reinstated content quadrupling.

The use of automated solutions to detect and remove terrorist content is also not straightforward. These solutions cannot replace consensus on what constitutes a terrorist organisation, and need to be informed by responsible terrorist designations from governments and intergovernmental organisations. It becomes even more complicated when harmful content originates from users that are not officially affiliated with terrorism or violent extremism, or when the content exists in a legal "grey area".

3. Leaving smaller platforms behind

Smaller platforms lack the resources necessary to deploy automated moderation tools at scale, which presents a dual risk. On the one hand, smaller platforms risk being left behind and penalised for not being able to comply with provisions where automated technology might be necessary. On the other hand, there is a risk of an uniformisation of the online moderation landscape and the expansion of what Evelyn Douek has labelled "content cartels", with smaller platforms turning to larger ones for content moderation tools (buying their services or replicating their moderation practices).

Tech Against Terrorism calls for greater support for smaller tech companies, in particular via development of data-driven moderation tools built with considerations for human rights and transparency on counterterrorism efforts, such as the Terrorist Content Analytics Platform. The development of these tools should be adapted to the needs of smaller platforms and respect their autonomy. Governments and larger platforms should support the development of these tools and facilitate their accessibility to smaller platforms, in respect of accountability and transparency.
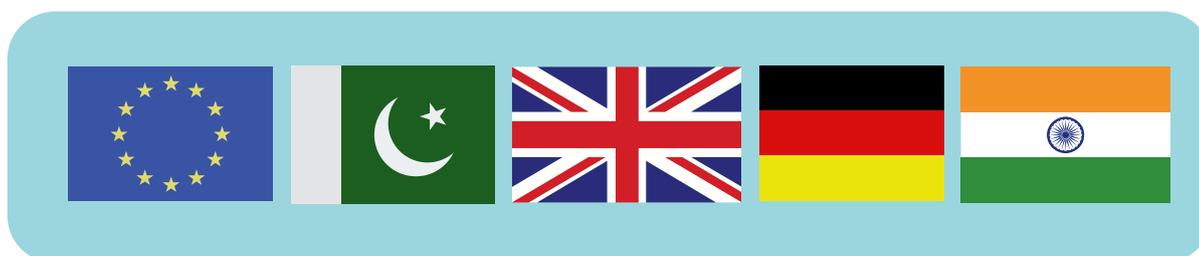
# 4. Delegation of legal adjudication to tech companies

One key emerging trend is regulation introducing mechanisms that outsource adjudication of illegality from courts and democratically-accountable institutions to private and unaccountable tech companies. This poses severe risks to the rule of law, not to mention that platforms, particularly smaller ones, do not have the legal expertise to adjudicate accurately, especially when they are facing time pressure such as removal deadlines.

International human rights standards underline that limits to freedom of expression should be decided by an independent judiciary body. However, by mandating tech companies to remove content at scale, many online regulations meant that are meant to counter online harms instead shift the responsibility of deciding what is harmful and/or illegal content to private tech companies. This is exemplified by the criticism made by David Kaye on the French "cyberhate law". The law itself did not create a new set of harms (it was based on restrictions to freedom of expression existing in French law); nevertheless Kaye underlined that "censorship measures", such as those implied by the duty to remove terrorist and hateful content, should not be delegated to private entities.

Tech Against Terrorism calls on governments to provide support mechanisms for tech companies to address terrorist and violent extremist content, and to provide clear legal bases to counter other online harms:

- With regard to terrorist content, we believe that designating terrorist groups in a responsible and accurate manner – whilst respecting human rights and freedom of expression – is an important tool that helps tech companies take appropriate action, whilst inscribing such action in the rule of law.
- Platforms should also be given sufficient time to properly assess the legality and harmfulness of content, and should be practically supported in this endeavour.

All laws that mandate platforms to remove flagged content within a short timeframe, or proactively remove certain types of content, are in effect placing the onus of adjudication of illegality on tech platforms. In our assessment, this includes:
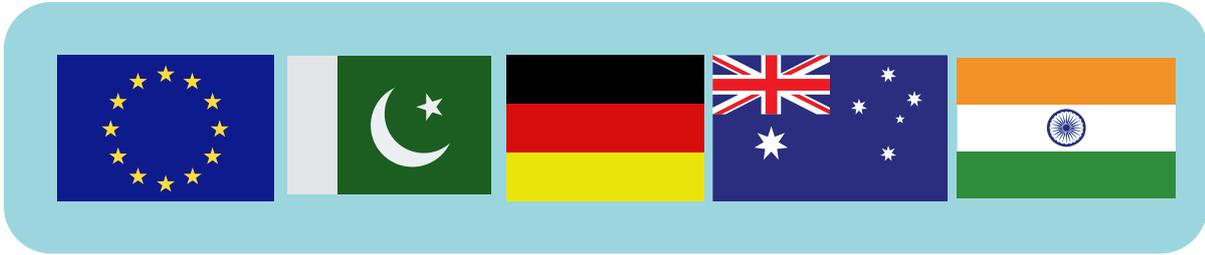
5. Holding platforms liable for user-generated content

The Online Safety Bill draft in the UK, and proposals to reform Section 230 in the US, suggest that platforms are increasingly likely to be held liable for user-generated content. Tech Against Terrorism cautions against holding platforms legally responsible for user content as this would heighten the risks for freedom of expression.

As the Global Network Initiative has warned, imposing liability on tech companies is likely to lead to the over-removal of content rather than tackling the underlying drivers of terrorist content on the Internet.

In addition, many platforms exist only as hosts or mere conduits. Forcing them to undertake moderation and content checks would open them to scrutiny and potential liability for third party content they have little to no oversight over.
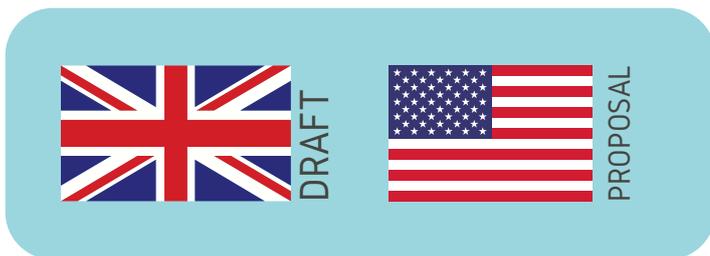
6. Placing legal liability on platform employees

Certain countries, including India and Pakistan, require tech companies to designate focal points for handling reports of violating content and user complaints. The UK draft Online Safety Bill takes this a step further by including a provision on "Senior Manager liability", which opens the way for senior managers to be held accountable for failing "to take all reasonable steps to prevent [an] offence being committed."

In some instances, employees of tech platforms have already been held legally liable for their companies' non-compliance with government requests. This goes beyond the usual fines that platforms can face for not abiding with regulations or government requests, with employees jailed or threatened with imprisonment in order to pressure platforms to comply.

Tech Against Terrorism warns against such provisions, which risk criminalising individuals engaged in countering the diffusion of terrorist and violent extremist material, rather than on those responsible for diffusing such content. In non-democratic countries with broad definitions of terrorist and harmful content, this further bears the risks of platforms and their employees becoming the targets of crackdowns on political dissent and non-violent speech.

Instead of holding platforms' employees responsible for terrorist content, there is a need to address the root causes of radicalisation and terrorism, and ensure that counterterrorism frameworks can be used to hold terrorists accountable for their online actions.

7. Local physical presence requirement

A number of regulations passed in 2020 and early 2021 require tech companies to establish a presence within the remit of a territorial jurisdiction – whether that be by appointing a focal point or nominating a legal representative or by establishing a physical office or by a data center within the country.

Complying with such requirements represents a significant challenge for smaller tech companies, especially as they are replicated throughout multiple countries. Potentially, small and micro-size platforms operated by 1-15 people will have to ensure a legal or physical presence in several countries if they wish to continue to operate there: a requirement that most smaller platforms will not be able to comply with due to the financial cost associated with it and will, as a result, be forced to stop their services in certain countries. Ultimately this is a threat to diversity and innovation in the tech sector.

Depending on the legislation and specific provisions, only larger tech companies have to comply with such requirements. However, these still present increased risks of governmental control over tech companies, such as via the legal liability of a platform's point of contact or users' data, as is the case with regulations mandating tech companies to set up data centres within a specific territorial jurisdiction. This risks having a country's authorities facilitated access to user data by diminishing the need to send complicated mutual legal assistance treaty requests across jurisdictions to access user data. Law enforcement and judicial authorities, including in non-democratic countries, can thus use such data centre requirements to facilitate information and content removal requests at the expense of users' privacy rights.

Given the global nature of the online space, Tech Against Terrorism warns against the multiplication of legal requirements forcing platforms to have a physical or legal presence in a country. Replicated across jurisdictions, this creates a multiplicity of impossible legal requirements.
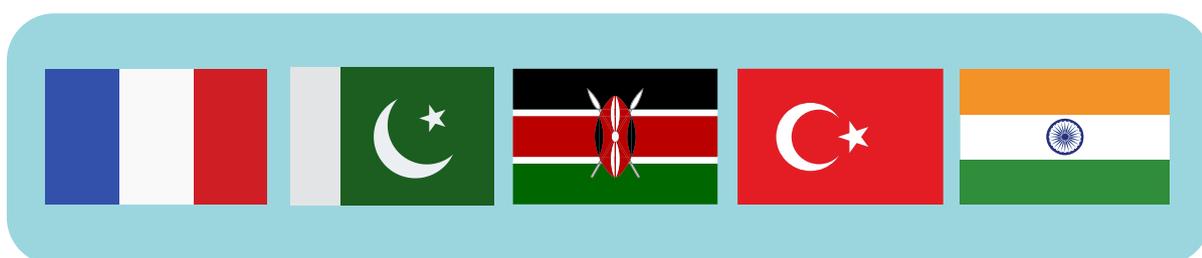
8. Mandating different requirements depending on platforms size

Some online regulations acknowledge that smaller platforms should not be expected to comply with the same level of demanding requirements than larger ones, and include provisions that only larger platforms need to comply with.

India and Turkey, for instance, include specific provisions for large platforms to comply with. However, the definitions or criteria used to define what constitutes a "large" platform are not always clear in these laws, and mandates further clarification from authorities in charge of overseeing the implementation of the laws. The Bill on Separatism in France also requires platforms over a certain user-base size in France to comply with specific requirements on countering the spread of "illegal and hateful content", including a review of their algorithms.

Tech Against Terrorism welcomes the consideration given to smaller platforms in certain laws and amendments. However, we recommend policymakers to clarify in the regulatory frameworks the categorisation of platform size and to consider not only the user-base but also platform resources (financial, human and technical) in their categorisation process. This would ensure that platforms that lack resources are not misclassified.
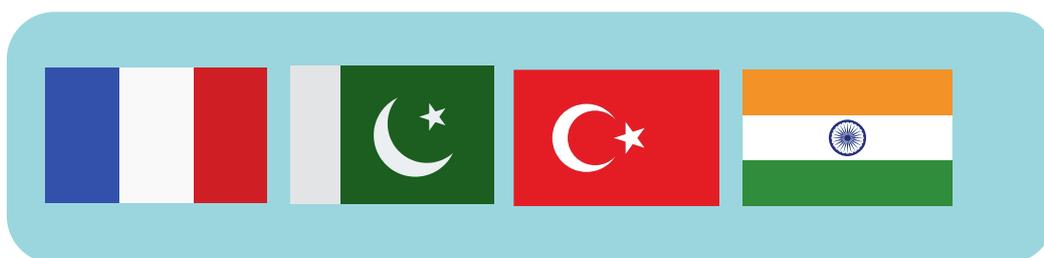
9. Transparency reporting expectations and requirements

Commendably the majority of online regulations introduced in 2019-2021 include provisions that seek to increase transparency and accountability from the tech sector.

Mandating detailed content standards

Some of the regulations analysed in this Handbook state that tech platforms should have clear and detailed content standards for users to understand what is allowed or not on the platform. In certain instances, regulations outline what should be included in the content standards, and mandate or recommend platforms to explicitly prohibit the types of content that are covered in the regulation itself.

The EU Regulation 2021/784 states that platforms should have a clear prohibition of terrorist content in their community guidelines, whereas the EU Digital Service Act and the UK Guidance for Video Sharing Platforms outline what platforms should raise in their content standards. The 2020 Rules in Pakistan and the 2021 Guidelines in India both go a step further and require platforms to add to their content standards the list of content prohibited in the laws.
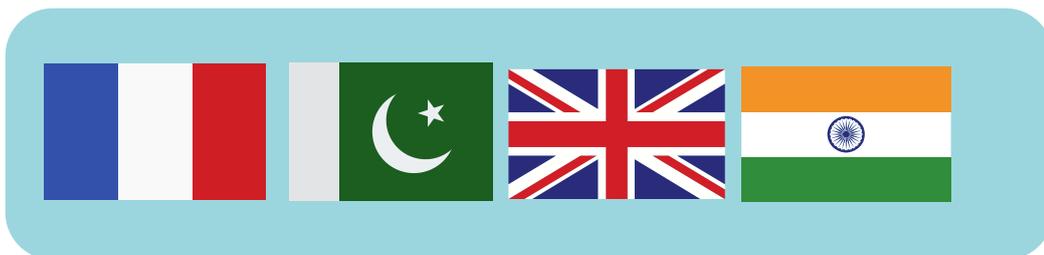
Increasing transparency reporting

On transparency, the proposed Online Safety Bill in the UK demands that platforms publish transparency reports on their compliance with the Bill. The EU Regulation 2021/784 on Addressing the dissemination of terrorist content online, will require tech companies to publish transparency reports on their efforts to comply with the regulation, and outlines metrics for transparency reporting by governments and competent authorities. France's "cyberhate" law also calls for increased transparency from both the tech and government sectors, and requires the country's audio-visual authority to publish an annual report on the enforcement of the law.

However, such calls for increased transparency and accountability are either formalising what is already best practice across the tech sector (e.g. clear and accessible guidelines); or risk setting unrealistic expectations as to what metrics should be included in transparency reports by not considering platform capacity, diversity, and functionality, and therefore applying a misguided one-size-fits-all approach to transparency.

Governments should support the tech sector's transparency efforts and, in doing so, acknowledge that standardised reporting practices will not lead to meaningful transparency, as different platform purposes, policies, and processes will inevitably lead to discrepancies in data points.

Tech Against Terrorism recommends governments to support our Guidelines for Transparency Reporting on online counterterrorism efforts. Our Guidelines focus on a small number of core metrics to facilitate evaluation of performance over time, and fully recognise the importance of platform diversity.

We also call for increased transparency from governments on their online counterterrorism efforts by supporting our transparency Guidelines for governments. These Guidelines are – just like the Guidelines for tech companies – meant to drive increased transparency around a small set of core principles to improve overall transparency from governments.
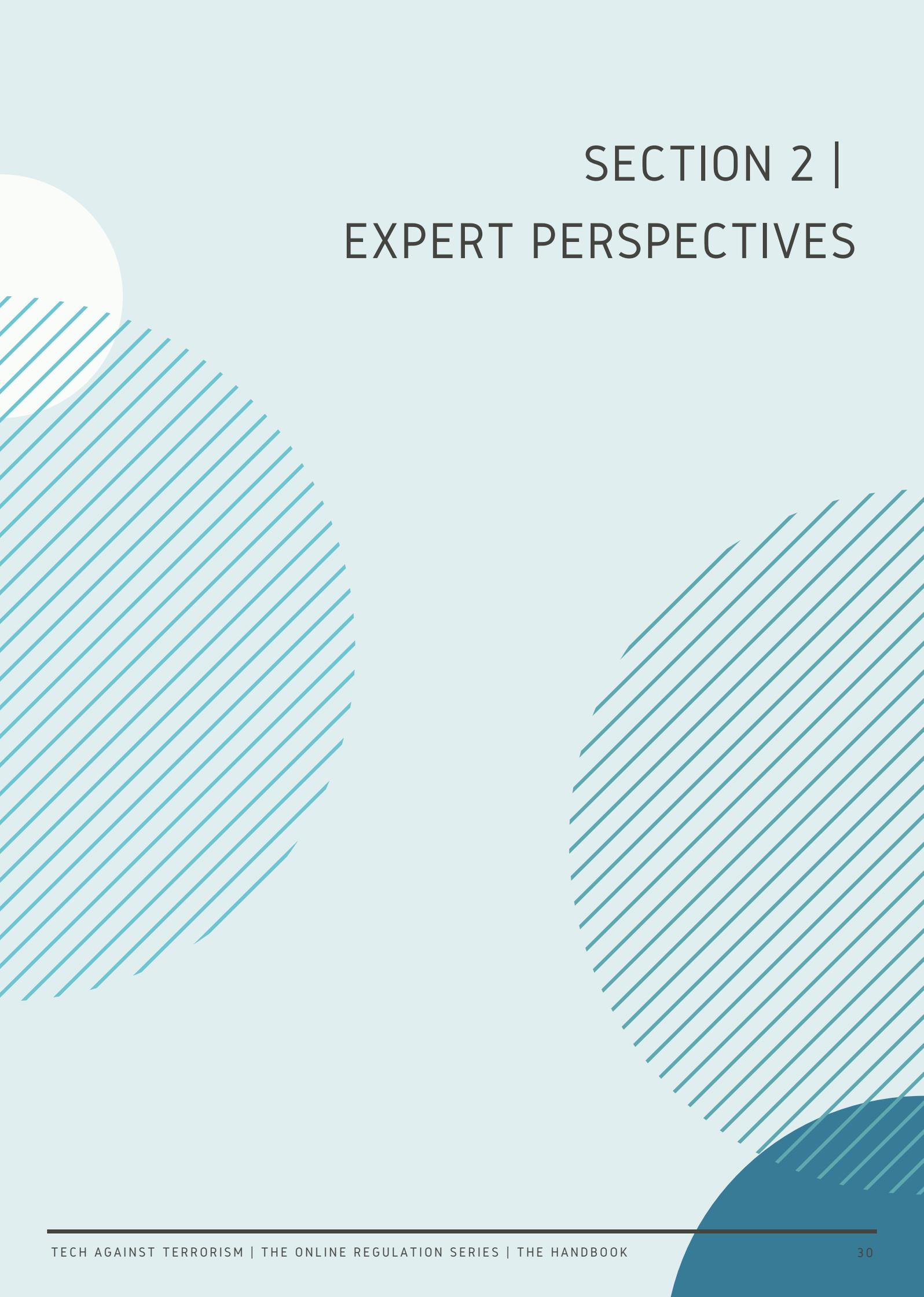
SECTION 2 |
EXPERT PERSPECTIVES

THE STATE OF ONLINE REGULATION | ACADEMIC ANALYSIS

As global regulation of online speech has increased, so has academic analysis of such regulation. Here, we provide an in-depth look at some of the key academic analysis of global regulatory efforts.

Key takeaways:

- Academics agree that global regulation of online speech has changed drastically over the past two decades. More recently, there has been a move away from governments allowing platforms to set the rules for online speech, to an increase in government led regulation, some of which has changed some of the ground rules of the modern internet.
- Generally, academics agree that improvements in regulation is needed in order 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 does not hold tech companies to account.

Background: evolution of content moderation

Academic research demonstrates that regulation of online speech has drastically evolved since the emergence of the internet. Whilst big tech companies initially had rudimentary moderation guidelines,[2] most now have intricate moderation policies and mechanisms . Since most global speech platforms were founded in the US, the online speech landscape has largely been shaped by the US First Amendment.[3] However, academics highlight that this is rapidly changing.

---

2. Both Facebook and YouTube initially had a one-page document to guide decision-making.
3. 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).

Jonathan Zittrain has 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.

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 should 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 "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 allowed platforms to act as de facto regulators of the wider industry. 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. In fact, several of the mechanisms that scholars noted to 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)[4] 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.

---

4. 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[5] 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. She warned that this leads to misguided policy proposals that risk having limited effects on actually tackling terrorism and terrorist use of the internet.

It is worth examining what patterns academics have identified across regulations introduced globally 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.

---

5.   Including in the European Union, the United Kingdom, France, Pakistan, and the Philippines.

Questioning of intermediary liability shields

Perhaps the 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.

Removal deadlines

Academics have noted an increase the introduction of removal deadlines in global regulation. Such deadlines compel companies to remove illegal or harmful content within a specified timeframe.[6] 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 have warned that 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."

---

6. In the proposed French law, it was 24 hours (one hour for terrorist and CSA material), in the proposed EU regulation it is one hour, and in Australia companies are compelled to remove content "expeditiously", with no specific timeframe).

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 because "harmful" is rarely precisely defined. Several categories of potentially "harmful" speech might be legal, and 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 Kate 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.

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. 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.[7] The way companies solve this is often by "geo-blocking" content in some jurisdictions, making it inaccessible for users in such countries, whilst allowing it in other jurisdictions. Increasingly, governments and courts have begun to compel companies to remove access to content violating national legislation in all global jurisdictions (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[8] 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.

---

7. Some early encounters of this challenge being content defaming the late Thai King Bhumibol, or the founder of Turkey, Mustafa Kemal Atatürk.
8. The most notable case being the United Kingdom.

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 companies 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.[9] Whilst, as Douek notes, the sheer scale and technical requirements might 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.

---

9. Germany's NetzDG law is one example.

THE FUTURE OF ONLINE REGULATION | EXPERTS' RECOMMENDATIONS

Here we 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 policymakers 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.

- <u>Prescriptive model:</u> Under this formulation, governments would set out clear rules and specify the proactive measure that platforms would be required to abide by, thereby 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 current practices. 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.

- <u>Flexible model:</u> 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. However, it 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 legally 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.

Potential US Communications Decency Act Section 230 reform

Following the Trump Administration's executive order in May 2020, which directed 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 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. Academic 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 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. The agency would also seek to make 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 offers a global structure (as opposed to national laws),and provides a framework for ensuring that both companies and governments comply 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 – is also applicable in 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.

This is further discussed in the next section on international human rights law as a blueprint for content moderation

Social media councils

Global 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 governments 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 should commit to executing council decisions.

This suggestion was supported by 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 appeal and review body, with a first version being launched on a national scale as a trial.

EXPERT PERSPECTIVES |

INTERNATIONAL HUMAN RIGHTS LAW AS A BLUEPRINT FOR CONTENT MODERATION: BENEFITS AND CHALLENGES

In this section we distil some of the leading academic analysis on how international human rights laws can be used to inform content moderation.

Key takeaways for tech companies

- International human rights law arguably provides the best global framework on which companies can base content moderation policies. International human rights law can help companies push back against repressive government regulation. Furthermore, the approach is not necessarily prescriptive, and allows a degree of tech platform autonomy in terms of what speech to prohibit. However, applying international human rights law as the foundation for content moderation is complex to implement at scale, and experts highlight that it might not always provide platforms with clarity on how to address thorny speech issues. To adequately implement this approach, companies would need to feel comfortable assessing specific speech cases against international human rights law, meaning that it might not necessarily be the best solution for smaller companies with limited resources.

What is international law and international human rights law?

International law refers to a set of international treaties and norms that define States' legal responsibilities to each other. There are various bodies of international law. The body most frequently mentioned with regards to content moderation is international human rights law (IHRL). IHRL consists of a number of international treaties, of which the Universal Declaration of Human Rights (UDHR) is the most fundamental.

- The UDHR, together with the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights form the so-called International Bill of Human Rights, and lay out principles for the rights of all people.

- In 2011, the <u>UN Human Rights Council endorsed the UN Guiding Principles</u> (UNGPs) on Business and Human Rights. The UNGPs are considered as part of the set of documents that constitute international human rights law.

All of these documents form the basis of the Tech Against Terrorism Pledge for smaller tech companies.

What parts of IHRL are relevant to content moderation?

There are three documents that are regularly mentioned by academics examining IHRL's applicability to content moderation:

- ICCPR, in particular Article 19 of the covenant: This Article ensures freedom of expression for everyone, but also sets out the cases in which expression might be restricted. Article 20 of the ICCPR mentions two specific types of speech (propaganda for war and advocacy of national, racial, or religious hatred) that should be prohibited.

International Covenant on Civil and Political Rights (ICCPR)

Article 19

1. Everyone shall have the right to hold opinions without interference.
2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
(a) For respect of the rights or reputations of others;
(b) For the protection of national security or of public order (ordre public), or of public health or morals.

Article 20

1. Any propaganda for war shall be prohibited by law.
2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.
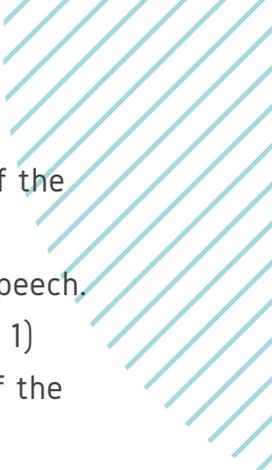
- The Rabat Plan of Action: The Rabat Plan was developed by the UN Office of the High Commissioner for Human Rights in 2013 to provide guidance on how to restrict content in line with Article 20 in a way that does not restrict free speech. It suggests a six-step test to assess whether speech should be criminalised: 1) context 2) speaker 3) intent 4) content and the form of speech 5) extent of the speech 6) likelihood and imminence.

- The UNGPs on Business and Human Rights: The UNGPs establish a framework that compels companies to a) avoid causing harm to human rights and mitigate such impact of their operations b) make high-level commitments to human rights c) conduct due diligence to identify and address human rights risks d) implement mitigation strategies to safeguard human rights e) continuously review their efforts to respect human rights f) provide remedy via "grievance mechanisms" in case of violation.

Tech companies that have publicly committed to human rights standards often refer to the above documents. Facebook and Twitter have both said that international human rights standards guide their decisions, with Facebook saying it deployed the Rabat Plan's six-step test in making its decision to remove former US President Donald Trump from the platform. The Facebook Oversight Board (discussed at length in the industry initiatives section) also refers to the above documents when analysing specific cases.

Why use IHRL in content moderation?

The inception to IHRL being discussed as a framework for online content moderation is the work of human rights lawyer and former UN Special Rapporteur on Freedom of Expression David Kaye. In a 2018 report to the UN's Human Rights Council, Kaye recommended that companies ground their content standards in IHRL. Kaye argued that IHRL supersedes national laws and company Terms of Service with regards to companies' responsibilities to a global user base.[10] His recommendation was generally supported by several civil society and activist groups, and instigated several academic analyses of IHRL's applicability in online content moderation, many of which are discussed below.

Benefits of using IHRL as a blueprint for content moderation

The benefit of using IHRL as a basis for content moderation is clear: it aligns companies with globally agreed human rights standards. However, several academics have highlighted other benefits.

Kaye outlines two: Firstly, grounding policies in IHRL will allow companies to push back against state pressure to censor content, since IHRL would supersede national, and potentially repressive legislation. Secondly, despite ensuring freedom of expression, IHRL provides guidance on what speech should be prohibited and on what grounds. To that end, Kaye argues that IHRL would give companies a globally recognised framework to design tools to deal with illegal content.

Several academics have further analysed the benefits of using IHRL in content moderation. Barrie Sander[11] and Susan Benesch,[12] agree that using IHRL creates a framework that disincentivises ad hoc implementation and enforcement of policies. Benesch and Michael Lwin[13] both agree that there is a clear benefit in creating a universal standard for content moderation, in a way that allows for flexibility with regards to content moderation decision-making.

---

10. Kaye David (2018), <u>A Human Rights Approach to Platform Content Regulation.</u>
11. Sander Barrie (2020), <u>Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation.</u>
12. Benesch Susan (2020), <u>But Facebook's Not a Country: How to Interpret Human Rights et Human Rights Law for Social Media Companies.</u>
13. Lwin Michael (2020), <u>Applying International Human Rights Law for Use by Facebook.</u>

This would particularly benefit larger tech companies due to their global reach. Benesch also notes that the IHRL approach also avoids the risk of larger companies imposing their values and speech norms on the rest of the world.

Experts have also highlighted that the IRHL is beneficial given how easily it can be operationalised. Sander, like Kaye, highlights that IHRL gives flexibility to tech companies, and that the framework does not necessarily dictate the outcome of a platform's decisions. Several academics also agree that IHRL can be adapted for a corporate context.

Sander, Benesch, and Evelyn Mary Aswad,[14] argue that platforms can use their Terms of Service and content standards to meet Article 19's requirements, that stipulates restrictions on speech must be "provided by law". The UN Human Rights Council has clarified that "provided by law" means it has been "made accessible to the public" and "formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly".

Aswad also argues that platforms can adapt their practices to meet the "necessity" as outlined in Article 19, by restricting enforcement against prohibited speech by using "the least intrusive means" possible. Lwin also stresses that the Rabat Plan to Action can be used by tech companies to facilitate decision-making in hate speech related cases, and encourage companies to use the 6-factor test mentioned above to design a scoring system to help decide on whether to allow or restrict certain speech.

Experts also point out that the UNGPs can improve tech company appeal and redress mechanisms. Lwin notes that the UNGPs outlines criteria for "non-judicial grievance mechanisms" (see below) which tech companies can utilise.

14. Aswad Evelyn Mary (2018), The Future of Freedom of Expression Online.

UNGP on Business & Human Rights

Article 31

In order to ensure their effectiveness, non-judicial grievance mechanisms, both State-based and non-State-based, should be:

(a) Legitimate: enabling trust from the stakeholder groups for whose use they are intended, and being accountable for the fair conduct of grievance processes;

(b) Accessible: being known to all stakeholder groups for whose use they are intended, and providing adequate assistance for those who may face particular barriers to access;

(c) Predictable: providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation;

(d) Equitable: seeking to ensure that aggrieved parties have reasonable access to sources of information, advice and expertise necessary to engage in a grievance process on fair, informed and respectful terms;

(e) Transparent: keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism's performance to build confidence in its effectiveness and meet any public interest at stake;

(f) Rights-compatible: ensuring that outcomes and remedies accord with internationally recognized human rights;

(g) A source of continuous learning: drawing on relevant measures to identify lessons for improving

Challenges in using IHRL as a blueprint for content moderation

Academics who are generally positive towards the IHRL-centred approach to content moderation do however list important caveats. Firstly, despite several benefits to the feasibility of IHRL in content moderation, it would still require adaptation to work for content moderation at scale. This is because the majority of IHRL was not created for the internet age and the speed and scale with which speech is transmitted online.

Further, it does not provide a safeguard against human or algorithmic error in content moderation. Researchers have also highlighted that platforms might not be equipped to accurately assess speech against IHRL standards.

Benesch caveats that platform content standards might not be clear enough to replace the "provided by law" requirement, and highlights the potential misuse of the restricted speech provisions in Article 19 (national security and "rights and reputation of others") – two areas which are often used as justification for draconian internet regulation. These restrictions, often labelled as the "legitimacy prong" of the Article 19 restrictions, are the most difficult adaptation for tech companies to make. This is both because it might allow for censorship and because it does not cover several of the content categories that platforms might restrict due to platform values or their business interests. Sander points out this conflict, noting that employing IHRL is unlikely to solve trade-offs between competing interests. Sander further argues that platforms might be resistant to comply with IHRL obligations if it threatens their commercial interests. Aswad agrees, and questions whether it is realistic to expect platforms "to refrain from restricting speech at the expense of their bottom lines [of maximising profit]?"[15]

---

15. Aswad Evelyn Mary (2018), <u>The Future of Freedom of Expression Online.</u>

Some scholars have directly questioned the applicability of IHRL to content moderation. Evelyn Douek[16] is sceptical as to whether IHRL will solve the fundamental challenges around content moderation. One of Douek's key criticisms is that IHRL does little to solve some of the thornier content moderation challenges and to account for jurisdictional and cultural differences. Douek highlights how the inherent flexibility that IHRL offers also means that it contains gaps and inconsistencies that leaves it subject to differing interpretations. For example, Douek argues that IHRL could provide justification for allowing Holocaust denial content, and that it does not account for the vastly different historical and cultural contexts that has led to such material being banned in specific jurisdictions, but not in others. Douek argues there is indeterminacy in IHRL that leaves decision-making to platforms, and questions whether all platforms have the competency to assess speech against IHRL standards.

Likewise, Brenda Dvoskin,[17] notes the challenge of accommodating for differing jurisdictional approaches and applications to online speech, and that two separate jurisdictions can – whilst ostensibly committed to IHRL – come very different conclusions about fundamental speech challenges. Dvoskin highlights the issue of upload filters, noting that whilst they have been encouraged by the Court of Justice of the European Union in the well-publicised Glawischnig-Piezcsek vs Facebook case, the use of such technology would be forbidden in the American Convention of Human Rights. Dvoskin argues that this is an example of how international standards can be contradictory rather than clarifying.

Lastly, Dvoskin and Douek both agree that there is a risk that platforms accepting IHRL as a standard will give them a veneer of legitimacy, but without actually forcing them to tackle the policies and practices of their platforms that risk harming online speech.

16. Douek Evelyn (2020), The Limits of International Law in Content Moderation, UCI Journal of International, Transnational, and Comparative Law (forthcoming 2021).
17. Brenda Dvoskin, "International Human Rights Law Is Not Enough to Fix Content Moderation's Legitimacy Crisis" (2020)

EXPERTS PERSPECTIVE | TECH SECTOR INITIATIVES

Although governments have passed legislations aimed at countering terrorist and harmful online content in recent years, content moderation in practice remains mostly a matter of by the tech sector.[18] This entails companies drafting and applying their own rules for moderating user-generated content on their platforms in line with their values, business interests, or when they voluntarily comply with industry standards legal enforcement.[19]

The predominance of self-regulation, 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 the impact on freedom of speech online. In this entry, 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.

---

18. Article19 (2019), Social Media Councils: Consultation.
19. The standards set by the Global Internet Forum to Counter Terrorism are one example of this.

The 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 tens of members and has taken a prominent role in the Christchurch Call to Action – launched following a far-right terrorist attack in March 2019 in Christchurch, New Zealand, which was livestreamed on Facebook. Tech Against Terrorism has been one of the GIFCT's core partners since its inception, organising its inaugural workshop in San Francisco in 2017.[20] Since then, Tech Against Terrorism has supported the GIFCT knowledge sharing programme by organising workshops and e-learning webinars. The Tech Against Terrorism's Mentorship Programme is also meant to assist tech 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.[21] 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 new members that meet "leadership criteria".

---

20. Tech Against Terrorism (2017), Tech Against Terrorism San Francisco Workshop and US Launch of the Global Forum to Counter Terrorism.
21. The Khalifa-Ihler Institute defines itself as a "global peace-building organization dedicated to building and empowering thriving and inclusive communities".

The GIFCT also runs the Hash-Sharing Consortium to help member companies moderate terrorist content on their platforms.[22] The consortium is a database of hashed terrorist content.[23] Members can add hashes of content they have previously identified as terrorist material to the database. All companies using it are able to automatically detect terrorist material on their platforms and prevent further uploads.

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.[24] 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.[25] The GIFCT said in its 2020 report 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%
- Radicalisation, Recruitment, Instruction: 2.1%
- Christchurch, New Zealand, attack and Content Incident Protocols (Christchurch, 6.8% Halle attack, 2% Glendale attack 0.1%)

---

22. The Consortium was set up prior to the GIFCT by the four founding companies in 2016.
23. Hashing technology allows for the attribution of a unique fingerprint tot a photo or audio content, thus facilitating its identification without having to see the content itself. This can be used to facilitate the identification of terrorist content and prevent its upload. As the GIFCT explains it: "An image or video is "hashed" in its raw form and is not linked to any original platform or user data. Hashes appear as a numerical representation of the original content and cannot be reverse-engineered to recreate the image and/or video. A platform needs to find a match with a given hash on their platform in order to see what the hash corresponds with."
 24. 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.
25. The reports also provide information about the Content Incident Protocol (CIP) and the URL Sharing mechanisms. Two others technical mechanisms implemented by the GIFCT to ensure the facilitated removal of terrorist content, grater collaboration between platforms, and limit the spread of terrorist content following an attack in the case of the CIP.

Academic and online regulation expert, Evelyn Douek,[26] 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.[27]
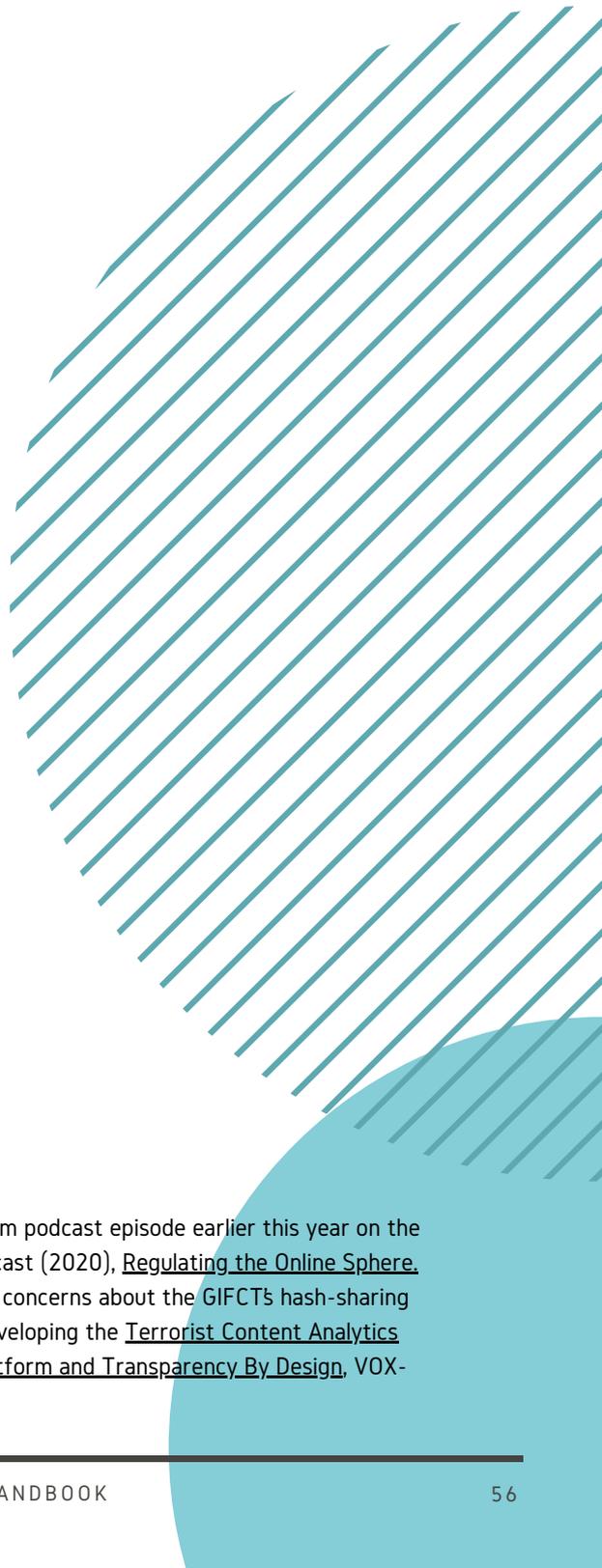
26. Evelyn Douek spoke about her criticism of the GIFCT in a Tech Against Terrorism podcast episode earlier this year on the complexities of regulating the online sphere. See: The Tech Against Terrorism Podcast (2020), Regulating the Online Sphere.

27. In November 2020, Tech Against Terrorism responded to an article mentioning concerns about the GIFCT's hash-sharing database, and how we are planning on taking into account these concerns when developing the Terrorist Content Analytics Platform. See: Tech Against Terrorism (2020), The Terrorist Content Analytics Platform and Transparency By Design, VOX-Pol.

The Facebook Oversight Board

Facebook announced in 2018 that it would set up an independent body 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.[28]

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 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.

---

28. Since then, the Oversight Board has already published 12 decisions, including three related to Facebook's dangerous individuals and organisations policy, and two related to violence and incitement. All decisions can be found here. The most high-profile decision made by the Oversight Board as of June 2021 is the decision to upheld Facebook's restriction on former President Donald Trump's Facebook and Instagram accounts. This decision was accompanied by a number of policy recommendations, which led Facebook to change its approach to "newsworthy content" and content made by "influential users", including politicians.

Whilst the concept of the Oversight Board has been welcomed, it has nonetheless drawn criticism. 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.

Klonick 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. Klonick described this initiative as "misleading", given that it would not hear any user appeals.

The 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 as well as experts in online safety (including cyberbullying) and 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,[29] Director of the Free Expression Project at the Centre for Democracy & Technology, and an expert on free expression online and intermediary liability.

---

29. At Tech Against Terrorism, we have previously welcomed Emma Llanso in our podcast and our webinar series:
Tech Against Terrorism (2020), Summary of our webinar on transparency reporting for smaller tech companies; and The Tech Against Terrorism Podcast (2019), How we fight terrorism while protecting human rights.

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 Centre, 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 will 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 experts in: tech policy, online safety, and young peoples' mental health, with the plan to grow to about 12 experts.

Following the announcement of this first Advisory Council, TikTok announced an Asia Pacific Safety Advisory Council in September 2020, as well as European Safety Advisory Council in March 2021.

SECTION 3 |
GLOBAL ONLINE REGULATION

SECTION 3 |
GLOBAL ONLINE REGULATION

ASIA-PACIFIC

Singapore

ASIA-PACIFIC | SINGAPORE
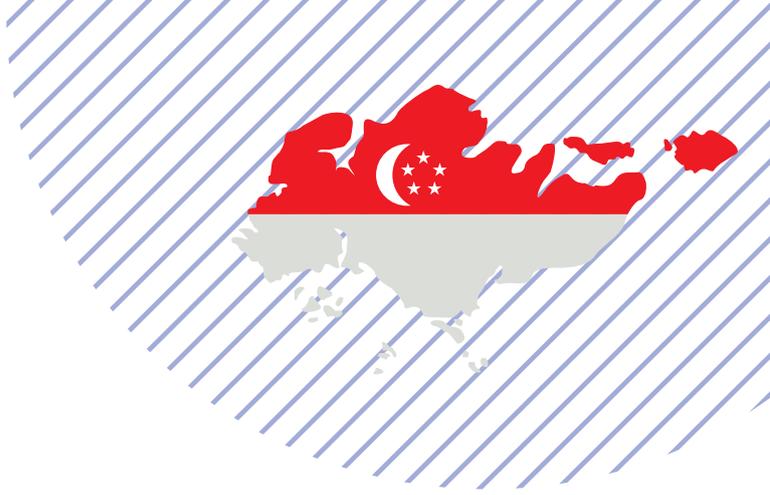
Singapore's online regulation framework does not follow any of the different global key trends identified by Tech Against Terrorism. However, Singapore is one of the few countries where online regulation also applies to encrypted messaging services.

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, last updated in December 2020, 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 aims to tackle the spread of misinformation through correction and removal orders.

Main bodies overseeing online regulation:

- Infocomm Media Development Authority (IMDA), the government agency regulating the information and communication technology as well as 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" 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."[30]

---

30. Around $744.105

TECH AGAINST TERRORISM COMMENTARY

## Lack of consideration for smaller platforms

Following the enactment of POFMA, major tech platforms – including Facebook, Twitter and Google Search – were granted a temporary exemption for certain requirements. However, only granting a temporary exemption to big tech platforms risks creating a two-tier regulatory system that penalises smaller platforms instead of supporting them. The exemption was granted to big platforms to give them time to adapt to the new regulation, which is commendable for acknowledging that tech platforms need an adaptation period to accommodate legislations that can significantly impact their moderation processes. However, it raises questions regarding the lack of consideration toward smaller tech platforms, which lack the resources and capacity required to adapt their processes to new regulations. Given the penalties that companies face for non-compliance, 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.

Concerns over potential breaking of end-to-end encryption

The regulation of encrypted messaging services (EMS) is often discussed on the basis of countering terrorist use of the internet and child sexual abuse However. Singapore is the first country to have passed a legislation regulating online content that also applies to EMS. It remains to be seen how the POFMA will be enforced for encrypted messaging services. Most EMS rely on end-to-end encryption (E2EEE), which provides a layer of protection making it technically impossible for anyone but the sender and recipient(s) to view the content of a message. Considering that third-parties cannot view the contents of E2EE communications, it is unclear how platforms can be expected to remove content or post corrections in compliance with the POFMA.

Applying the same regulatory framework for content shared on public-facing platforms (e.g., social media and content sharing platforms) raises significant questions with regard to freedom of expression and the right to privacy, as it mandates the monitoring of private communications to identify, correct, and remove "online falsehoods". Tech Against Terrorism cautions against government regulation that requires tech companies to modify their systems and processes that could weaken encryption.

ASIA-PACIFIC | PAKISTAN
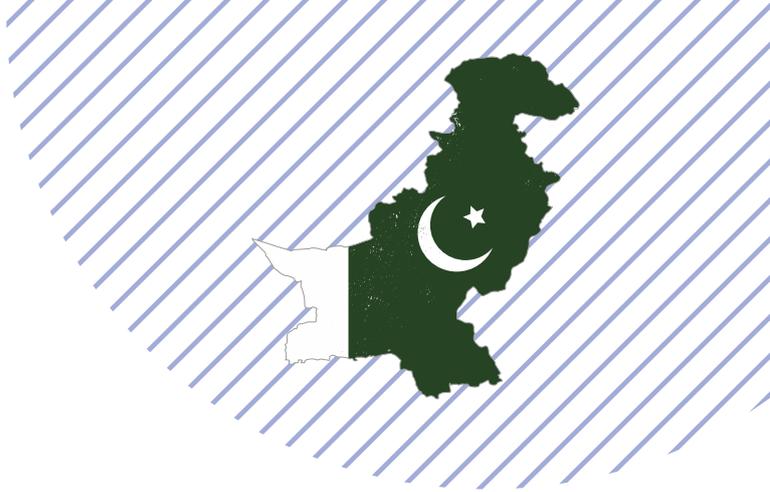
Amongst the different global key trends identified by Tech Against Terrorism, Pakistan follows:

- Mandating different requirements depending on platform size
- Mandating short removal deadlines
- Encouraging increased reliance on automated moderation tools
- Mandating a focal point for user complaints or law enforcement
- Mandating a local presence

Since 2016 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 legal framework for counterterrorism in Pakistan. The ATA does not specifically cover 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 incites hatred or "gives projection" to a terrorist actor.

Pakistan's regulatory framework

- Anti-terrorism Act, August 1997, sets the framework for Pakistan's counterterrorism response.
- Prevention of Electronic Crimes Act (PECA), August 2016, provides a "comprehensive legal framework" to counter electronic crimes and related investigations. Section 37 of the PECA notably covers what content is accepted or not in the country.
- Section 37 of the PECA, The Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules 2020, outlines, in vague terms, which content should be blocked in the country. In June 2021, the Ministry of Information and Technology announced modified draft rules under the PECA, which replicate most of the provisions of the 2020 Citizen Protection (Against Online Harm) Rules.[31]
- Citizen Protection (Against Online Harm) Rules, October 2020, aims to regulate online content, including terrorist material and hate speech. Passed with immediate effect on 20 October 2020. The Rules have received criticism from the tech sector and civil society organisations:
- In a statement published in November 2020, the Asia Internet Coalition criticised the lack of public and stakeholder consultation around the Rules, and argued that its members "will be unable to operate in the country with this law in place."
- The Rules have been contested by a coalition of civil society organisations, who argue that the Rules contradict some rights guaranteed by the constitution, including freedom of expression. They submitted a petition to the Islamabad High Court, which heard it in January 2021. Following this, the Court requested the government to amend the Rules by 2 April 2021, and asked Pakistan's Attorney General to submit a report in this regard. In response to this request, the Pakistani Prime Minister announced on 29 March 2021 the creation of an inter-ministerial committee to review the social media rules.

Relevant national bodies

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

---

31. The modified draft is to re

Key takeaways for tech platforms

- Via the PECA and 2020 Citizen Protection Rules, Pakistan explicitly prohibits terrorist use of the internet, and the sharing of terrorist content on social media.
- Under the PECA, individuals posting terrorist material online can be held liable and face jail terms.

2020 Citizen Protection Rules[32]

- The 2020 Rules introduced a new framework mandating tech platforms to remove and block access to content, including for content published by Pakistani citizens outside of the country's jurisdiction, that are against:

  - "Glory of Islam"
  - "Integrity, security and defence of Pakistan" – this applies to terrorist related content
  - "Public order" – this includes fake or false information that threatens public order
  - "Decency and morality"

- With regard to terrorist, extremist content, hate speech, and incitement to violence, service providers and social media platforms must deploy mechanisms preventing the upload and livestream of such content.

- Under the 2020 Rules, platforms will have to:

  - Remove content within 24 hours following removal requests issued by the Telecommunication Authority, and within 6 hours in "case of emergency". The Rules allow for the removal to be differed for up to one month if needed to support a criminal investigation.[33]

---

32. In June 2021, Business Recorder reported that the 2020 Rules were repealed and to be replaced by the Modified Removal and Blocking of Unlawful Online Content Rules under PECA. However, the modified draft Modified Removal and Blocking of Unlawful Online Content Rules replicate most of the requirements that were included in the 2020 Rules.
See: Amin Tahir (2021), Procedure, Oversight and Safeguards 'Removal & Blocking of Unlawful Online Content Rules, 2021' modified, Business Recorder.
33. As of June 2021, it is unclear whether the modified Removal and Blocking of Unlawful Online Content Rules will maintain the same timeframe for content removal. However, according to the modified draft rules, platforms will have 48h to submit an explanation as to why they did not comply with an order.

- Include a list of prohibited content covered by the law in their Community Guidelines, and implement the necessary processes to identify such content.
- Provide "any information or data or content" in a decrypted, readable, and comprehensible format to the Investigation Agency.

- In addition, "significant social media" companies are to:

  - Register with the Authority within three months after the Rules coming into effect.
  - Establish a permanent office in the country, preferably in Islamabad, within 6 months of the Rules coming into effect.
  - Appoint a compliance officer based in Pakistan within three months of the regulation coming into effect.
  - Appoint a "grievance officer" based in Pakistan to receive users' complaints and respond to these complaints within 7 days of receipt. The "grievance officer" should be appointed within three months of the regulation coming into effect.
  - Name and contact details of the "grievance officer" should be made publicly available to users.
  - Use "suitable content moderation" system, including Artificial-Intelligence based tools, and hire content moderators "well verse with the local laws".[34]
  - Set up one or more database centres in the country, and store user data within territorial boundaries, within a year of the regulation coming into effect.[35]

- The Telecommunication Authority will establish a web-based complaint mechanism (web-based) for anyone to request the removal or blocking of content.

- Companies that fail to abide by the 2020 Rules can be blocked from operating in the country or face consequential fines of up to approximately US$ 3,236,246.[36]

---

34. This was added with the modified draft Modified Removal and Blocking of Unlawful Online Content Rules in June 2021.
35. As of June 2021, it is unclear whether the modified draft Modified Removal and Blocking of Unlawful Online Content Rules will maintain this provision.
36. 500 million rupees.

TECH AGAINST TERRORISM COMMENTARY

24h removal deadline and mandated upload filters

- The 2020 Rules introduce a stringent regulatory framework for tech companies operating in Pakistan, particularly with regard to deploying tools to proactively detect and remove content and the 24-hour removal deadline. Since the limited timeframe would make in-depth examination of any flagged content difficult, this framework risks making tech companies over-remove content in order to ensure compliance rather than risk fines.

- By requesting tech companies to prevent the upload and livestream of terrorist and extremist content, the Rules effectively mandate the use of upload filters by tech companies. Such filters have been criticised by digital rights advocacy organisations over the risks of removing legitimate content.

Physical presence in the country

- The Rules require tech companies to have a physical presence in the country, as well as local data centres to store Pakistani user-data within the country's territorial boundaries. This presents risks for user privacy as the it would allow the government and law enforcement to have facilitated access to data by no longer needing to go through Mutual Legal Assistance Treaties (MLAT) to request platforms to disclose information about their users.

Designating a complaint officer

- The designation of a complaint officer and the requirement for contact details to be public present risks of abuse by designating one individual as the sole responsible for content moderation on a platform. This is problematic as it is unclear whether that person would be legally liable if their platform does not introduce the necessary moderation measures requested by the Rules. It also creates an unnecessary burden for tech companies, which can otherwise manage user complaints via online forms rather than through a designated compliant officer.

ASIA-PACIFIC | THE PHILIPPINES
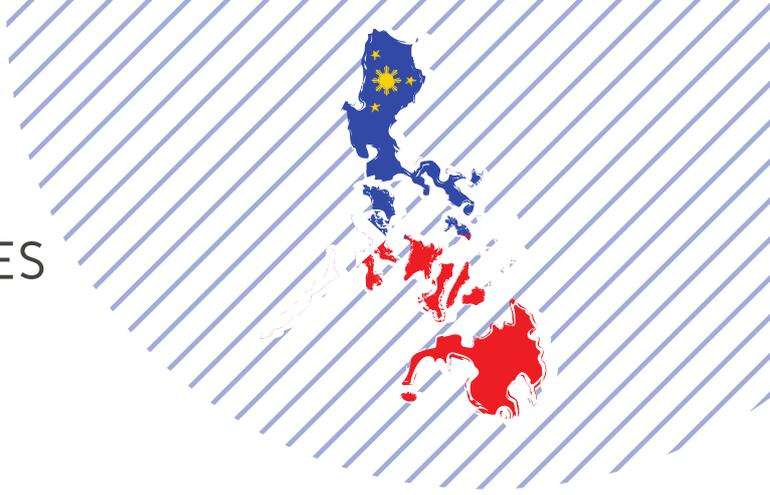
The Philippines is one of the countries worst affected by terrorism in the world. In 2020, the country ranked 10th in terms of numbers of terrorist attacks according to the Global Terrorist Index.[37] The country has long been investing in counterterrorism, and there have been some signs that the government might introduce legislation that targets online terrorist content. The country has a growing internet penetration rate and increased use of social media (+8.6% in 2019-2020).

Philippines' regulatory framework

- Anti-Terrorism Act (ATA), July 2020, provides the legal framework for the country's counterterrorism response.
- The Act has been contested with petitioners, including legal experts and human rights advocates, arguing that it contains unconstitutional provisions.
- The Philippines' Supreme Court will hear the oral arguments of the petitioners in 2021.
- Cybercrime Prevention Act (CPA), September 2012, is the country's regulatory framework for information and communication technologies.

Key takeaways for tech platforms

- Tech companies are currently exempt from liability for user-generated content posted on their platforms.
- Recent suggestions to expand the Anti-terrorism Act of 2020 to allow for the regulation of social media, indicates that tech platforms could become liable for online terrorist content in the future.

---

37. Institute for Economics & Peace (2020), Global Terrorism Index 2020: Measuring the Impact of Terrorism.

ASIA-PACIFIC | AUSTRALIA

Amongst the different global key trends identified by Tech Against Terrorism, Australia follows:

- Mandating short removal deadlines
- Outsourcing legal adjudication to tech companies
- Mandating transparency and accountability

The Broadcasting Services Amendment (Online Services) Act of 1999 has been regulating harmful and illegal online content in Australia since the late 1990s. The Act 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 of 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.

- The <u>Online Safety Bill</u>, passed in June 2021, sets out to reform and expand existing online safety regulations. The Bill introduces five schemes to deal with different types of harmful online material:
- Four schemes already exist in law, but are being updated by the Bill. These schemes include: cyber-bullying, image-based abuse, online content.
- One is new: the adult cyber abuse scheme.
- The Bill also includes new, shorter, takedown deadlines as well as industry codes.
- The <u>Online Safety Charter</u>, 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 <u>report</u> on how the government and tech industry could improve their ability to prevent and respond to future online crisis events. As a result of the report's recommendations, ISPs and the government have agreed to a new <u>protocol</u> to allow the blocking of websites hosting graphic material depicting a terrorist act or violent crime.
- Australia is a signatory of the <u>Christchurch Call to Action</u>.

Relevant national bodies

- The <u>e-Safety Commissioner</u> is empowered under the Enhancing Online Safety Act 2015.
  - The Commissioner administers the Online Content Scheme, and can issue notices to service providers over content that violates the Criminal Code Amendment Act 2019.
  - The Commissioner can tell ISPs to block access to material that exposes people in Australia to online terrorist and extreme violent material, but only during crisis events.
  - The Online Safety Bill enables the Commissioner to utilise a new rapid website blocking power to block websites hosting abhorrent violent or terrorist material during an online crisis even, such as the Christchurch attack in 2019. The Bill also requires search engines and app stores to remove access to a website or app that "systematically ignores" take down notices for class 1 material, such as child sexual abuse material.
  - The Commissioner produces <u>annual reports</u> on their performance, including on their assistance and investigations.

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 <u>Criminal Code Amendment Act 2019</u> –– can be sanctioned by:
    - A fine of around $1.5 mn[38] or up to three years in prison (for an individual providing the content services or hosting services).
    - A fine up to around $7.5mn[39] or 10% of annual revenue for each offense (for a company).

- Examples of potential violations of the Criminal Code Amendment Act of 2019 include:
    - Providing a content service or hosting service which can be used to access abhorrent violent material.
    - Failing to ensure expeditious removal or cease hosting of it following notification from authorities.
    - Failing to refer details to the Australian Federal Police after becoming aware of such content being available on their service.

- The e-Safety Commissioner can initiate investigations relating to online content and is able to enforce actions like issuing notices:
    - The Commissioner can block access in Australia to certain content hosted overseas, by notifying the Australian ISPs about 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 "<u>reckless</u>" about its service hosting abhorrent violent material.

- The Online Safety Bill introduces the following:
    - 24-hour deadline for Online Service Providers when receiving a notice from the eSafety Commissioner for image-based abuse, cyber-abuse, cyber-bullying, and seriously harmful content.
    - Expanded cyber-bullying scheme for children, which enables the removal of material from online services including social media platforms, games, websites, messaging and hosting services.

---

38. AU$2.1 mn
39. AU$10.5 mn

- Basic online expectations to establish mandatory reporting requirements that will allow the eSafety Commissioner to require online services to provide specific information about online harms. This could include information about responses to terrorism and abhorrent violent material, or volumetric attacks. Services will have to respond on how they will uphold these expectations and they can be penalised if they fail to report.
- An update to Australia's Online Content Scheme. This reflects and simplifies the current regime in Schedules 5 and 7 of the BSA, with some clarifications of material and providers of services captured by the scheme, and extends the eSafety Commissioner's take-down powers for some material to international services in some circumstances. This includes bodies and associations that represent sections of the online industry may develop industry codes.
- A new cyber abuse scheme allows the eSafety Commissioner to remove seriously harmful abuse online when websites, social media and other online services do not remove content after a complaint is made.
- These protections will be backed by civil penalties for service providers who fail to comply.
- Extended powers for the eSafety Commissioner:
  - eSafety Commissioner has a new rapid website blocking power. This can be used to block websites hosting abhorrent violent or terrorist material during an online crisis event.
  - eSafety Commissioner can require search engines and app stores to remove access to a website or app that "systematically ignores" take down notices for class 1 material under the online content scheme, such as child sexual abuse material.

TECH AGAINST TERRORISM COMMENTARY[40]
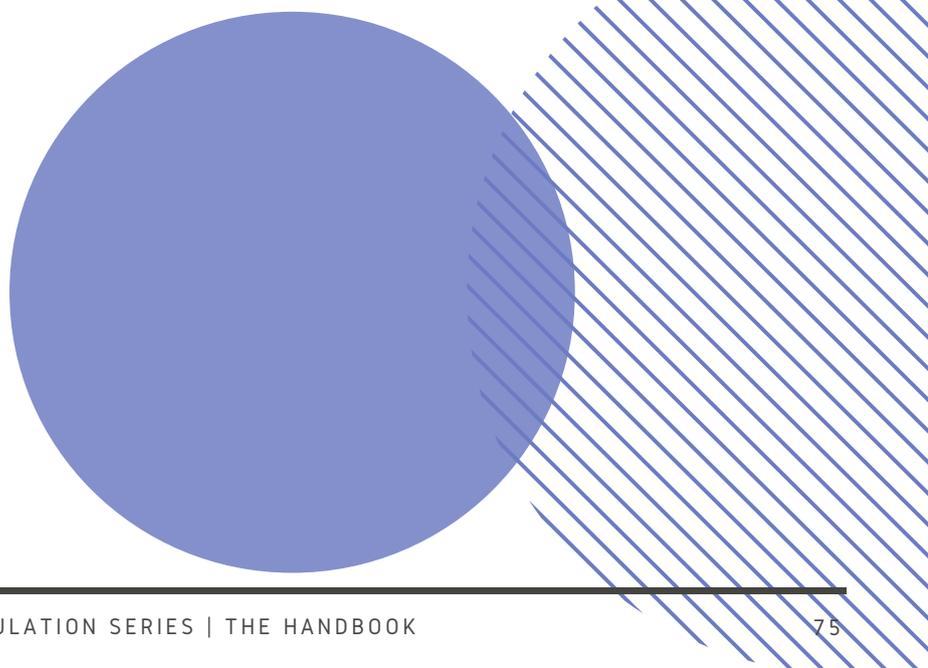
e-Safety Commissioner

- It is commendable that the main body in charge of coordinating and encouraging action from tech companies, the e-Safety Commissioner, has a clear legal standing. This ensures that several of the instruments provided (such as removal orders) are carried out in accordance with the rule of law. It is also positive that the e-Safety Commissioner produces annual reports on their performance, including on their assistance and investigations.

Rule of Law

- The Online Safety Bill risks leading to extensive takedown of legal (but 'harmful') speech. For example, whilst cyber bullying and abuse are issues that tech companies should counter for ethical reasons, compelling them to do so under threat of potential liability and financial penalties risks undermining the rule of law. Whilst some aspects of bullying and abuse are anchored in Australia's criminal code, the definitions provided in the Act suggest that the law will potentially lead to removal of large amounts of legally allowed speech. In a democracy, speech that is legal offline should not be illegal in the online space. If harms need countering online, they should be prohibited in law before legislation is created to remove such content from the internet.

- The tech sector should not develop codes that can subsequently be introduced into law with legal liability and subsequent financial penalties. Whilst improved industry codes should be encouraged, it is important that legislation is determined by democratically accountable institutions. Thus, there are some concerns regarding the legality of the development of industry codes – which may be developed by bodies and associations that represent sections of the online industry – within the Online Safety Bill.

---

40. The below comments can also be found in our submission and recommendations to the Online Safety Bill's consultation, see here.

Freedom of Expression

- We are concerned the Online Safety Bill has no clear references to safeguards that prevent the erroneous removal of content as a result of blocking or removal requests. This is particularly serious for link deletion and app removal requests, as these are severe steps with a potentially detrimental impact to freedom of information if carried out over extensively. Furthermore, there is no reference to redress mechanisms in the Bill.

- There are a number of imprecise definitions that we believe will negatively impact freedom of expression. The definitions provided for child cyber bullying and cyber abuse seem to build on a perceived 'common sense' approach as opposed to legal concepts. This therefore risk decisions being assessed subjectively. Not only could this lead to the removal of legal content, but it will also be difficult to operationalise for tech companies.

- We have some concerns around the Abhorrent Violent Material (AVM) scheme. Whilst the scope of the law is clear, we worry that imprecise definitions of "terrorist act" and calls for companies to remove content "expeditiously" could encourage tech platforms to remove content that is shared with the purpose of documenting terrorist offences and war crimes. Such content can serve as crucial evidence in court proceedings. We appreciate the necessity to restrict access to content that risks becoming viral in the immediate aftermath of a terrorist attack. However, due to the drastic measures that the Bill allows for, the Government should ensure that there are sufficient safeguards in place in case of wrongful blocking and that appropriate redress mechanisms are identified.

Smaller tech companies and tech sector capacity

- The Online Safety Bill does not explicitly refer to smaller tech companies, which often do not have the capacity to take swift action due to limited staff numbers or subject matter expertise on various harm areas. Since it is well-established that terrorists predominantly exploit smaller platforms for exactly this reason,[41] it is disappointing this is not reflected in the Act. Specifically, we worry that instruments such as the removal and blocking deadlines of 24 hours (which are punishable by steep fines) will severely harm competition and innovation.

Lack of Consultation

- The Criminal Code amendment law passed through both houses of parliament in a remarkably short time. Similarly, the Online Safety Bill entered parliament only 10 days after the public consultation on the Bill closed. This limits the possibility of consultation from the industry or civil society, or for policymakers to amend draft legislations in time to incorporate recommendations submitted during a consultation process.

- Tech Against Terrorism participated in the consultation for the Online Safety Bill. To read our full submission and recommendations, see here.

---

41. To read more about this, please see our analysis on ISIS's use of smaller platforms and the DWeb to share terrorist content – April 2019 here.

ASIA-PACIFIC | INDIA
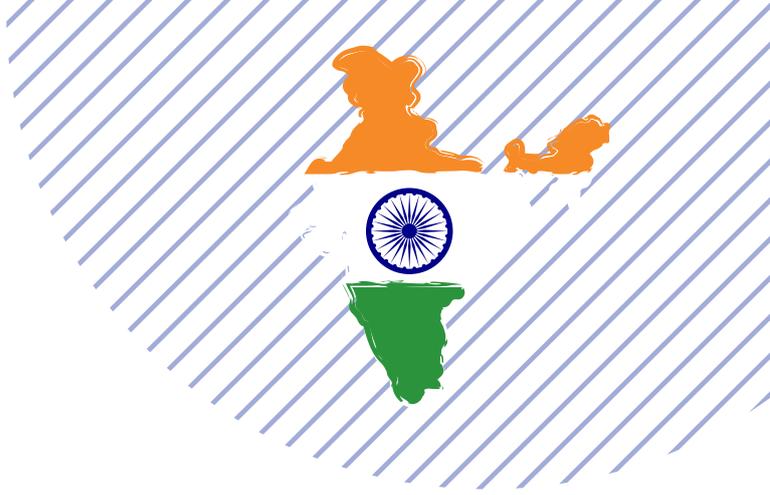
Amongst the different global key trends identified by Tech Against Terrorism, India follows:

- Mandating different requirements depending on platform size
- Mandating short removal deadlines
- Encouraging increased reliance on automated moderation tools
- Mandating a focal point for user complaints or law enforcement
- Mandating a local presence
- Mandating transparency and accountability

With almost 500 million internet users, and several high-profile cases of online disinformation contributing to offline violence, content moderation is a pressing issue in India. Regulation of content is covered by different legislations under the Indian Penal Code, the Information Technology Act (ITA), and Criminal Procedure Code. In February 2021, a new regulatory framework dedicated to online content was passed, the Guidelines for Intermediaries and Digital Media Ethics Code Rules.

Terrorist use of the internet is mainly regulated through cybercrime laws, covered by Section 66F of the Information Technology Act, which regulates cybercrimes and electronic commerce. The 2021 Rules also regulate content that, amongst others, "threatens the unity, integrity, defence, security or Sovereignty of India".

India's regulatory framework

- The Information Technology Act (ITA), <u>passed</u> in June 2000 and <u>amended</u> in 2005, is the framework for regulating cybercrime, including cyberterrorism, in the country.
- <u>Shreya Singhal v. Union of India</u> (2012), a landmark decision by the Indian Supreme Court in 2015, absolves tech companies from having to actively monitoring their platforms for illegal content.
- <u>Guidelines for Intermediaries and Digital Media Ethics Code Rules</u>, passed in February 2021, formalised what online content is prohibited in the country, and allows Indian authorities to request content removal. Tech companies were given three months to comply with the 2021 Guidelines, from when the law came into effect on 25 May 2021.

Key takeaways for tech companies

Tech platforms operating in India are exempt 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. The compliance guidelines were also listed under the Information Technology (Intermediary Guidelines) Rules, 2011; from 2021, platforms are to follow the requirements laid out in the new 2021 Guidelines.

ITA, Section 69A:

- Tech platforms can be asked to remove or block access to certain content deemed to threaten the sovereignty, integrity, and public order of India. Non-compliance can be penalised by jail terms and fines.
- This requirement has been re-asserted in the 2021 Guidelines, which require service providers to prohibit certain content from their services.

2021 Guidelines:

- Under the 2021 Guidelines, platforms are to
  - Ensure that their services cannot be used to share illegal content, including content that "threatens the unity, integrity, defence, security or Sovereignty of India".
  - Explicitly cover the content prohibited in 2021 Guidelines content in their policies to user.
  - Remove content prohibited by law, or disable access to it, within 36 hours after being notified by the Indian authorities (whether by court order or notification by a government agency).

- The 2021 Guidelines create a duty of due diligence for tech companies, which will have to:
    - Appoint a Grievance Officer and make the Officer's contact detail public, as well as detail how users can submit complaints. The Grievance Officer needs to be an Indian national residing in the country, and complaints have to be acknowledged within three days and responded to within a month.
    - Make their policies easily accessible for users, and publish updates to their terms and policies.
- The 2021 Guidelines lay out additional due diligence requirements for larger platforms, labelled as "Significant social media intermediaries(s)" in the regulation:
    - Appoint a Chief Compliance Officer to oversee compliance with the guidelines and to be held liable if platforms fail to ensure compliance.
    - Appoint a point of contact to maintain continuous coordination with law enforcement, and ensure that notifications and requests are swiftly responded to.
    - Appoint a Resident Grievance Officer.
    - Publish "compliance" reports every six months, detailing the complaints received and the content removed in response or as a result of proactive monitoring.
    - Establish an office in India, and publicly disclose its address.
    - Notify users of content removal and explain why content was removed.
    - Ensure that a redress mechanism is available for users to contest a removal decision.
    - Deploy automated tools and mechanisms to proactively identify and remove child sexual abuse and rape material, or content that had previously been removed.
- "Significant social media" that primarily provides messaging services are required to enable tracing of the original sender of a message.
    - This traceability requirement is limited to certain investigatory or prosecution purposes, including threats to national security and CSAM.
    - The Guidelines state that this should not be a requirement to disclose the content of a message.
- All of the above due diligence guidelines can apply to other intermediaries following a notification by the Central Government.
- The 2021 Guidelines also require tech companies to provide assistance to authorised government agencies conducting "investigative or protective or cyber security activities", and to provide information within 72h.

TECH AGAINST TERRORISM COMMENTARY

Introducing removal requests

The 2021 Guidelines significantly shift content regulation in India by codifying what type of content tech companies must prohibit on their services. Prior to this, the Indian government mostly relied on content takedown and blocking requests, which led the country to be the leading one in terms of number government removal requests sent to tech companies.

Under the new rules, responding to removal requests becomes more stringent, as companies have to remove or block access to content within 36 hours. The Guidelines further place the onus of policing online content to tech companies, and requires tech companies to specifically prohibit certain content that would have previously been the subject of a removal request.

A differentiated regime for large platforms

The 2021 Guidelines explicitly lay out additional due diligence requirements for "significant social media". This caveat is commendable for acknowledging that not all platforms have the resources and capacity needed to comply with extensive due diligence requirements. However, to ensure clarity on the scope of application, the Indian government should provide details as to what platforms are considered "significant", as it remains unclear.

Requirements for transparency and accountability

A positive aspect of the 2021 Guidelines is the emphasis on accountability and redress towards users. The different due diligence provisions require tech companies to make their policies easily accessible to users, and to be transparent about any removed content. However, caution is needed when mandating tech platforms to publish transparency reports. A one-size-fits-all approach to content moderation will not ensure more meaningful transparency, but could present significant challenge for platforms to collect the necessary data if the moderation policies and processes needed to underpin the production of a report are not in place.

Traceability requirements

A core element of the draft 2018 Guidelines, the traceability requirements is also included in the 2021 Guidelines, despite the criticisms raised against it since 2018.

Tech companies and digital rights experts[42] have criticised the traceability requirement, already present in the draft 2018 Guidelines.[43] Critics have in particular flagged the risks posed to end-to-end encryption, and the burdensome technical changes needed to comply with this requirement for platforms that do not collect metadata. Tech Against Terrorism is concerned that Mandating companies to track encrypted messages risks compromising existing encryption protocols and traceability assurances. This presents adversarial risks to tech platforms and their users, whose security and privacy could be comprised.

---

42. See: Newton Casey (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; Software Freedom Law Center (2019), Any regulation of online speech in India must safeguard the rights to free speech and privacy, Scroll.in; Wagner Kurt (2019), WhatsApp is at risk in India. So are free speech and encryption, Vox, 19 February 2019
43. In May 2021, WhatsApp filed a legal complaint contesting the 2021 Guidelines and specifically the traceability requirement arguing that it violates the privacy rights protected by the Indian Constitution.
See: Financial Post (2021), WhatsApp sues India govt, says new media rules mean end to privacy -sources.
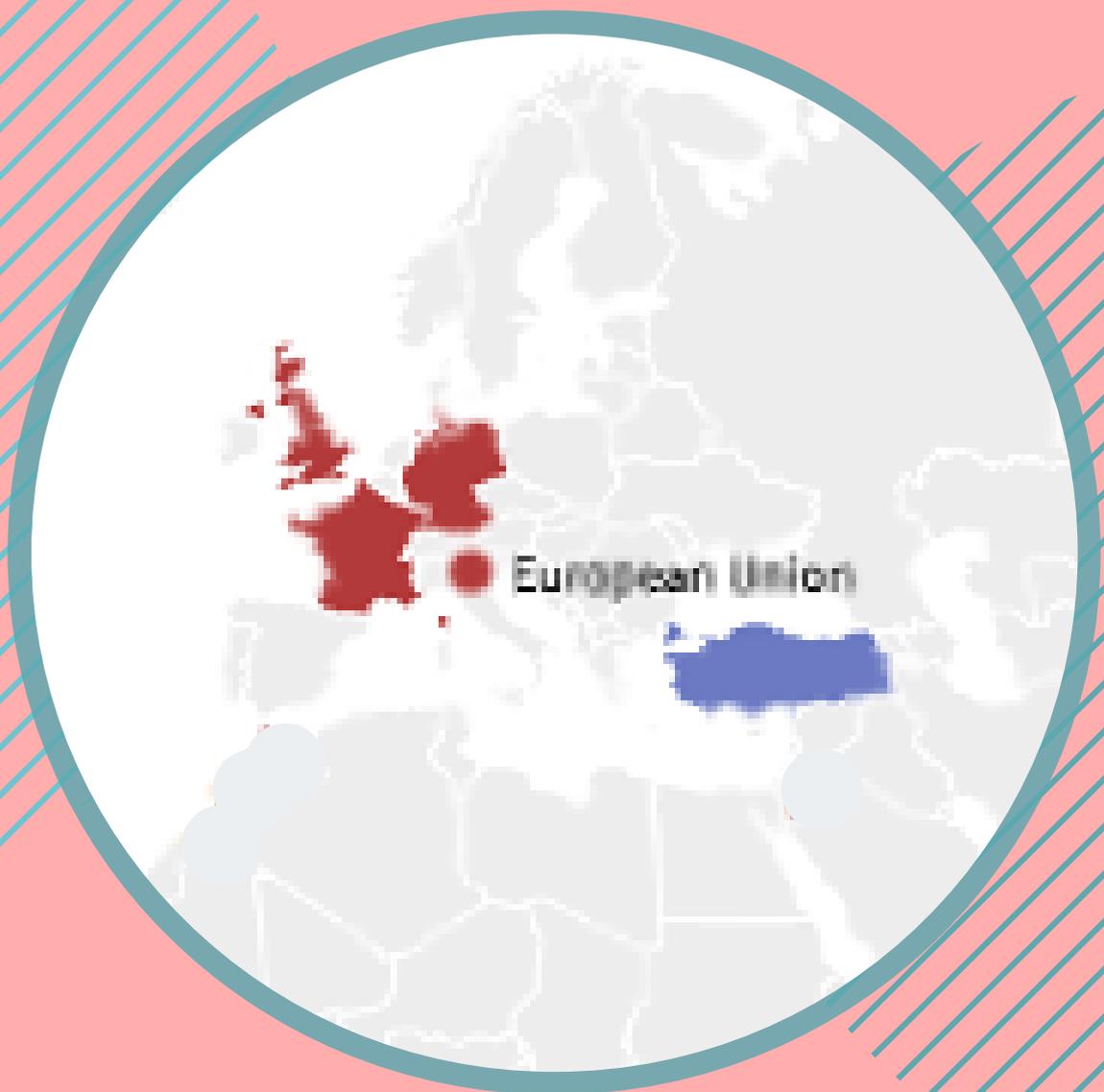
EUROPE | FRANCE
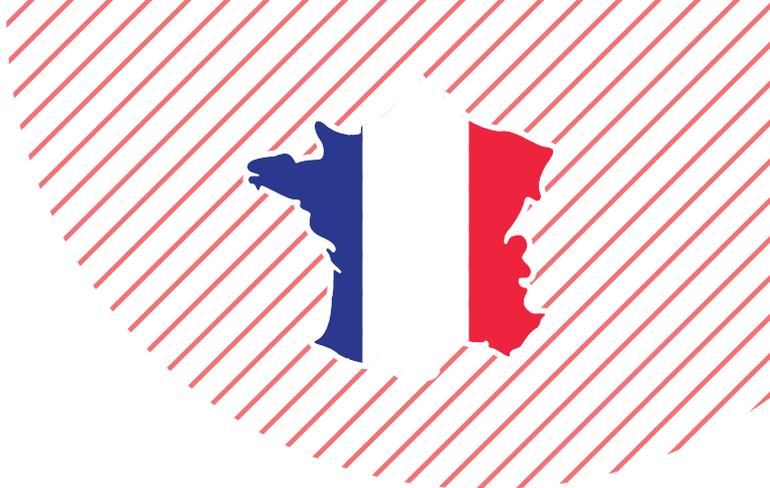
Amongst the different global key trends identified by Tech Against Terrorism, France follows:

- Mandating a local presence
- Mandating transparency and accountability

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 made tackling terrorist use of the internet a key pillar of its counterterrorism policy.[44] In line with this, the government had supported 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. The terrorist threat landscape in France, as well as recent attacks in 2020, have motivated changes to the existing counterterrorism and online content regulation frameworks, often meant at addressing the spread of online terrorist content and "online hate" in general.

France's regulatory framework:

- Countering online hate law. Adopted in May-June 2020, the so-called "cyber-hate" or "Avia" law,[45] 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 initially mandated companies to remove terrorist and CSA content within one hour of being notified by French authorities, and within 24 hours of being alerted[46] for other hateful and discriminatory content. Following a "censuring" by the French Constitutional Council, which deemed the law led to disproportionate risks to freedom of expression, the removal requirement was lifted and the law is now reduced to its preventive component.

---

44. Ministère de l'Europe et des Affaires Etrangères (2019), Terrorisme : l'action internationale de la France.
45. 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.
46. Including via user reports.

- Law on strengthening the provisions relating to the fight against terrorism, November 2014, strengthens France's counterterrorism approach and introduces the penalisation of "apologie du terrorisme" (terrorism apology or legitimation)[47] and incitement, including for content shared online.
- 2021 Amendment to law on confidence in the digital economy.[48] This law allows French judicial authorities to require a site to be blocked for hosting illegal content, including terrorist content. The 2021 amendment targets the circumvention of blocking orders – for example through the use of mirroring sites.
- France is a signatory and co-initiator of the Christchurch Call to Action.

Upcoming legislations:

- Endorsement of Respect for the Principles of the Republic and Counter Separatism Bill.[49] Commonly known as the "Bill against separatism", this law was presented by the government as a key-pillar of its strategy to counter Islamist radicalisation and terrorism. The law was approved after the first reading by the French parliament and senate and, as of June 2021, is to be going through another reading before being passed.[50] The original draft bill did not cover online content, however this was changed following the murder of Samuel Paty in October 2017.[51] The bill:

  - Penalises the malicious sharing of personal information online, that endangers the life of others. This article of the law is known as the "Samuel Paty" article.[52]
  - Penalises those who directly incite, legitimise or praise terrorism with a 7-year jail sentence and up to EUR 100,000 in fines. This applies to content shared on messaging platforms.
  - Punishes individuals who deliberately seek to circumvent moderation techniques used to counter and delete banned content.

---

47. In France the term "terrorism apology" is used to designate those promoting or praising a terrorist attack. This is similar to the condemnation of praise and legitimisation of terrorism acts and constitutes one of the key tenets of France's legislative framework against terrorist speech and content. See: Service-public.fr (last updated 2020), Apologie du terrorisme - Provocation au terrorisme.
48. Originally adopted in June 2004.
49. "Projet de loi confortant le respect des principes de la République et de lutte contre le séparatisme"
50. Following the first reading, the law was reviewed by a Commission Mixte Paritaire, a committee equally representing the Senate and the Assembly tasked with finding a compromise in case of persistent disagreement between the two legislative bodies. The law is expected to be approved once this review is completed.
51. Paty's murder was preceded by an online harassment campaign, which is being considered in the criminal investigation. See: Ouest France (2020), Assassinat de Samuel Paty. Suspect, gardes à vue, note du renseignement… Où en est l'enquête ?; and Devillier Nathalie (2020), Lynchage de Samuel Paty sur les réseaux sociaux : comment réguler les algorithmes de la haine ?, The Conversation.
52. Boukhelifa Florine (2021), Loi contre le "séparatisme" : les députés adoptent l'article dit "Samuel Paty", RTL.

- Creates new obligations for tech platforms, notably with regard to disclosing information about their algorithms and content moderation process.

- 2021 Counterterrorism and intelligence law. The law was approved by parliament at the first reading,[53] and incorporates elements of the emergency laws implemented following the violent Islamist terror attacks in 2015. The law notably incorporates the possibility for law enforcement to conduct "algorithmic analysis" of connection data and URLs provided by telecommunication operators.[54]

Relevant national bodies

- Superior Audiovisual Council ("Conseil Supérieur 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. The "Bill on Separatism" also grants the CSA powers to regulate content moderation.
- Ministry of Interior, which oversees – alongside judicial authorities – reports of terrorism apology and incitement, including for online content. Subsidiary bodies include"
- Pharos, France's online content reporting platform.[55]
- Cybercrime unit, which takes part in the coordination of content reported via Pharos and liaise with Europol's Internet Referral Unit.[56]

---

53. As of May 2021, the law has yet to be approved by the Senate before being promulgated.

54. The provision on algorithmics analysis has been criticised by digital right groups, including the Quadrature du Net, for its broadness. Critics argue that the lack of a specified scope of application could open the way for mass surveillance of internet connections and usage in France. In October 2020, the Court of the Justice of the European Union had already cautioned against France's practice to require user connection data to be kept for a year for counterterrorism purposes. The French Conseil DEtat later re-asserted that this provision was justified for counterterrorism purposes.
See: Jannic-Cherbonnel Fabien (2021), Projet de loi contre le terrorisme : cinq questions sur la surveillance par algorithme, une technique de renseignement critique, France Info; Le Monde (2020), La justice de lUE s'oppose a la colelcte a Cour de justice de l'Union européenne; 01Net.com (2021), Le Conseil dEtat approuve la conservation des données de connexion... en posant quelques limites.

55. Pharos can be used by any internet user to report online content via an online form. This is similar to user-reporting forms commonly found on online platforms to report content, with Pharos users choosing between different categories of illegal content – including one for terrorism (threat or legitimation). Users are then requested to provide additional information, including the URL, the type of platform on which the content was located and the time (optional). Reports can be anonymous.

56. "Office central de lutte contre la criminalité liée aux technologies de lInformation et de la communication »

- Online Hate Prosecutor Office:[57] Set up as the "judicial arm" of Pharos at the Paris Public Prosecutor Office in February 2021. This office monitors reports of hateful and extremist content made on Pharos, and prosecutes the user responsible for posting the flagged content. The office is also tasked to act as a representative of the French judiciary for tech companies, for example by organising meetings to inform tech companies of French law. The Prosecutor Office is one of the few provisions of the original "cyber-hate" law that was not censured by the French Constitutional Council in June 2020.
- State Secretary for Digital Affairs, which coordinates France's digital policy and matters related to the online regulatory framework.
- Digital Ambassador, coordinates international digital policy and transformation issues, including cyber security and online regulation.

---

57. Pôle judiciaire spécialisée contre la haine en ligne

Key takeaways for tech platforms

- France does not currently mandate tech platforms to regulate online content or hold platforms liable for failing to do so, despite the government's recent attempts to change this, such as with the introduction of the "cyber-hate" law.
- However, certain content categories are illegal under French law, including terrorist (incitement and apology) content, and actions can already be taken against this type of content:
    - French authorities can require a website to be blocked or a piece of content to be removed if terrorist content is located.
    - The 2021 Amendment to the law on confidence in the digital economy, expands the obligation to comply with a blocking order to any party that can contribute to a website being live.[58] It also grants authorities the possibility to require that a website or piece of content is removed from search engine results in France.
    - Individuals posting terrorist content risks seven years' imprisonment and around $119.400 fine.[59]
- The Bill on Separatism creates new obligations for tech companies, which will have to:
    - Have an established representative in France.
    - Review and respond to all content reports in a timely manner.[60]
    - Submit their algorithms for review to the CSA and disclose the number of moderators dedicated to content posted in France. This requirement only applies to platforms with a significant number of users in France.[61]

---

58. The amendments effectively expanding it from hosting services and internet access providers to virtually all internet service providers (e.g., web registrars, browsers, domain name providers).

59. i.e. EUR 100,000 – 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.

60. No details are provided as to what constitutes a "timely fashion" in the draft law.

61. The draft law does not provide more details on what constitutes a large platform, only referring to platforms that are the most important in number of connections.

TECH AGAINST TERRORISM COMMENTARY

Lack of practical tools to increase cooperation with the tech sector

The final version of the law on countering online hate, following the ruling of the Constitutional Council, promotes tools for cooperation and information sharing between platforms. However, it does so without specifying what these tools should be. More information on practical tools for cooperation and how the French government plans to support said cooperation, whilst also properly considering the impact on smaller platforms, would be more desirable.

Transparency and accountability requirements

The "cyber-hate" law calls for increased transparency and accountability from tech platforms and stipulates that the CSA publishes an annual report on its enforcement of the law. However, the transparency requirements do not take platform size and capacity into account. These criteria need to be considered in order to avoid unrealistic expectations as to what metrics platforms should include in their transparency reports.

The Bill on Separatism's current proposal to review platforms' algorithms is broad and does not specify how algorithms are to be reviewed by the CSA, nor what criteria will serve as the basis for assessment. Similarly, for the provisions on disclosing the number of moderators assigned to the French market, there is a lack of clarity regarding what specific number of moderators, or ratio of moderators per users, will be considered a sufficient threshold by the French government. At the time of writing, it is difficult to comprehend what these mandatory reviews are meant to achieve.

This risks penalising smaller platforms that do not have the resources to have moderators, or tools, dedicated to content published in France. Furthermore, there is a lack of clarity as to what this means for content originally published abroad but widely re-shared in France, and the potential risks for extra-territorial enforcement of national legislation.

Countering content moderation bypassing techniques

The Bill on Separatism and the 2021 amendment to the law on confidence in the digital economy include provisions to counter the use of content moderation circumvention techniques. Especially, when such techniques are used to disseminate terrorist content. This is sensible and commendable given that terrorists often deploy specific techniques to bypass content moderation. However, more cooperation with and practical support for tech companies, especially smaller platforms, are needed to help them improve their mitigations strategies in face of content moderation circumvention techniques, including mirroring sites.

EUROPE | GERMANY

Amongst the different global key trends identified by Tech Against Terrorism, Germany follows:

- Mandating short removal deadlines
- Encouraging increased reliance on automated moderation tools
- Mandating transparency and accountability

Germany has an extensive framework for regulating online content, <u>particularly with regards to hate speech and violent extremist and terrorist material</u>. Experts note that Germany's regulatory framework has helped set the standard for European online regulation.

Germany's Regulatory Framework:
- <u>The Network Enforcement Act (NetzDG)</u>, June 2017, aims to counter 22 different online offences, including cyberbullying, disinformation, child sexual exploitation, defamation, and terrorist use of the internet.
- <u>The Repair Act</u>,[62] April 2021
    - This includes <u>the February 2020 amendment</u>[63] to the NetzDG. The amendment aimed to counter right-wing extremism but was found unconstitutional. However, the Repair Act changed the amendment to ensure the law is constitutional.
    - The passed April 2021 Bill includes the changed amendment, as well as changes relevant to five other laws including: the Criminal Code, the law on Criminal Procedure, the NetzDG, the Telecommunications Act, the law of the Federal Criminal Police Authority.
- <u>The April 2020 amendment</u>, develops the requirements placed on tech companies in the NetzDG and adopts the obligations set in the European Union's <u>Audio-Visual Media Services Directive</u> (AVMSD) 2018 into national law.[64]

---

62. Gesetz zur Anpassung der Regelungen über die Bestandsdatenauskunft vom 30.03.2021, BGBl I, 448.
63. The Gesetzentwurf zur Bekämpfung des Rechtsextremismus und der Hasskriminalität.
64. 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.

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).[65]

Relevant national 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 reported content. Only social networks that are members of the FSM are able to use this mechanism.

Key takeaways for tech platforms

- NetzDG
  - As a general rule, the German government conscripts tech companies to carry out the requirements of the legislation set by the German government.
  - 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 criminal (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 must remove "manifestly unlawful content" within 24 hours and other prohibited content within 7 days.
    - Produce bi-annual transparency reports detailing how they respond to user reports.
    - Pay fines of up to around $2.4 mn[66] when failing to comply with the regulation.

- The April 2021 Repair Act adds further requirements to the NetzDG by compelling companies to:
  - Report user information (including the IP address of the user) to the Federal Criminal Police Authority, when they express certain criminal expressions online.

---

65. 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.
66. i.e. EUR 2 mn

- Assess whether their users express any of the prohibited types of expression and anything that would silence other users by intimidation. The list of expressions are included in 3a (2) of the NetzDG. They include:
    - Child pornography
    - Dissemination of propaganda and symbols from anti-constitutional organisations
    - Preparation of violent action against the state
    - Education and support of criminal or terrorist associations
    - Incitement to hatred
    - Representation of violence[67]
- For severe criminal actions such as the creation of a terrorist group, the police can also request the password of a user account, which a tech company must then provide.
- The transparency requirements compel 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.
- Strengthen appeal processes to allow users to challenge content removal decisions through a case-by-case review process.

---
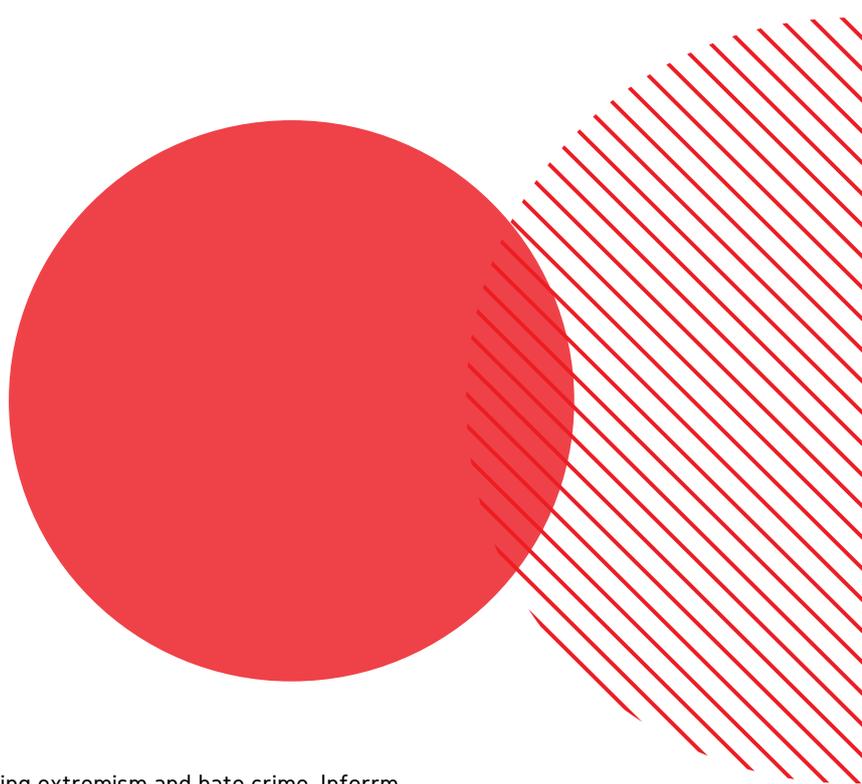
67. Bayer Judith (2021), <u>Germany: New law against right-wing extremism and hate crime</u>, Inforrm.

TECH AGAINST TERRORISM COMMENTARY

Responsibility of adjudication

The 2021 Repair Act says that offences that need to be reported on by tech companies to the police should be "identifiable at first sight" including content that related to terrorist and incitement to hatred offences. However, identifying what online content constitutes an offence is more complicated than the Repair Acts implies. This is particularly the case for "grey area content", which often blurs the line between terrorist or extremist material and legal content.[68] Furthermore, we deem that adjudicating on what is legal speech, both offline and online, should be done by governments, rather than private ones.

This responsibility should not only apply to the adjudication of terrorist content, but also for content that supposedly intimidates others into silence. Whilst we encourage governments to uphold freedom of speech, we deem that the current Repair Act does not give tech companies appropriate guidance on how to assess whether freedom of speech is limited. We would argue that this is a public responsibility, which should be guided by judicial or governmental institutions.

The rule of law

With particular regard to far-right extremist and terrorist content and the new amendment, governments should accurately designate far-right terrorist groups, as this provides tech companies with the legal grounding to remove related content from their platforms. We welcome Germany's leadership in this area, as the government has to date banned over 60 far-right violent extremist and terrorist organisations. This has as a result provided tech companies in Germany the appropriate legal grounding to moderate their platforms effectively.

---

68. For example, a piece of content that is produced in support of a terrorist or violent extremist group, but which does not make this support explicit nor directly call for violence. Or content that can be considered legal but "harmful".

Tight removal deadlines risks violating freedom of speech

Given that tech companies need to identify terrorist content "at a first glance", this is also reflected in the tight removal deadlines imposed by the law. Tech companies have 24 hours to remove "manifestly unlawful content". We deem this to be unwise. Among our concerns regarding risks to freedom of speech, we mirror digital rights organisations and activists[69] in cautioning that tight removal deadlines, in combination with high fines for platforms who are unable to moderate their platforms, might make companies err on the side of removal. This significantly undermines human rights, particularly freedom of speech. Furthermore, there is a risk of platforms reporting users to the police when they assess users' content and behaviour to be an offence at first glance, without properly considering the context (for example journalistic or educational) of the content. This would also risk the right to privacy. We therefore argue that rather than leaving the responsibility of adjudication to tech companies , governments should support tech companies in building expertise and tools to correctly identify "criminal expressions online" rather than set unreasonable expectations in legislative frameworks.

---

69. e.g. Article 19 and Daphne Keller

Law enforcement access to private users' information

We recognise the government adopted the Repair Act to ensure the constitutionality of the February 2020 amendment, and to increase the threshold for tech companies to report users'information to the police. However, we highlight the need for judicial oversight when handing over private users information to the police.[70] This is to ensure that human rights, particularly the right to privacy, is upheld when countering terrorist and extremist use of the internet. We understand the need for governments to require passwords of users for certain investigations, but this should not require tech companies to modify their technologies in order to provide passwords to law enforcement. Police requests for user information and data should not weaken security and privacy for all users of a particular service. This could happen if tech companies were required to change their encryption protocols and overall systems in order to provide said data.[71] Tech Against Terrorism welcomes any transparency the German government can provide, including requests for user data, in particular passwords, as well as their recurrence and whether these requests yields results for investigations.

Risk to smaller tech companies

Smaller tech companies face restricted capacity and often have fewer resources. They are therefore less likely to be able to adhere to these removal deadlines and reporting requirements, in particular for offences "identifiable at first sight". This risks the existence of smaller tech companies and to large tech companies with more resources monopolising online spaces and undermine the internet's diversity.

Concerns over negative global impact

The NetzDG has been used as a template for regulatory frameworks in other countries, despite 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.

Democratic countries should be conscious of the potential influence of their own regulations globally and consider how these regulations could be used as a template in non-democratic countries – with the risks this poses to freedom of speech and digital rights globally.

---

70. And not intelligence agencies as per the previous amendment.
71. On the question of law enforcement requesting tech companies to modify their technologies and encryption protocols to provide user information for a criminal or terrorist investigation, and the risks this present for all internet user, see: Apple (2016), A Message to Our Customers; Grossman Lev (2016), Inside Apple CEO Tim Cook's Fight With the FBI, The Time; Kahney Leander (2019), The FBI Wanted a Back Door to the iPhone. Tim Cook Said No, Wired; Electronic Frontier Foundation (2016), EFF to Support Apple in Encryption Battle.

EUROPE | THE EUROPEAN UNION

Amongst the different global key trends identified by Tech Against Terrorism, the European Union follows:

- Mandating short removal deadlines
- Outsourcing legal adjudication to tech companies
- Holding platform employees legally liable or demanding a focal point
- Mandating a local presence
- Mandating transparency and accountability

The European Union (EU) is an influential voice in the global debate on the regulation of online speech. For that reason, the EU's regulatory frameworks 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

- <u>European Counter Terrorism Strategy,</u> adopted in November 2005, which sets out the EU's priorities on countering terrorism in the Union.
- <u>European Agenda on Security,</u> 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
- <u>Directive (EU) 2017/541 on combating terrorism</u>, adopted in March 2017, and the key EU legal act on terrorism.[72]
- <u>E-Commerce Directive</u>, 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.
- <u>Audio Visual Media Services Directive</u>, 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.

---

72. 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: <u>https://europa.eu/european-union/law/legal-acts_en</u>

- Regulation 2021/784 on addressing the dissemination of terrorist content online, adopted in 2021, which compels companies to remove terrorist content within one hour and to introduce a range of measures to prevent terrorist content spreading on their platforms. The law will apply from June 2022.

Proposed regulation

- Digital Services Act (DSA), unveiled in 2020 and seeking to impose new rules to combat illegal and harmful content online. This includes an obligation to introduce a user reporting mechanism, measures examining tech company algorithms, and transparency reporting. Large online platforms will need to carry out risk assessments and undergo audits of their content moderation systems. The proposal also suggests slightly tweaking the liability scheme for tech platforms. Our response to the initial DSA proposal can be found here.
- Proposal for laying down harmonised rules on the use of artificial intelligence (AI), unveiled in 2021, and seeking to increase transparency and accountability in AI systems and regulate the use of "high-risk" AI.

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

Platforms are currently exempt from liability for user generated content, but this will change when the regulation on addressing the dissemination of terrorist content online starts applying in June 2022.

- Regulation on addressing the dissemination of terrorist content online (from June 2022)
    - One-hour removal deadline (Article 3). Companies are to remove content within one hour of receiving a removal order from a "competent authority" (which each Member State will be able to appoint – more information below). If the platform shows "systematic and persistent" failure to meet the one-hour deadline, it could result in penalty fees of up 4% of the company's global annual turnover.
    - "Specific measures". Companies are to introduce "specific measures" to prevent terrorist content if instructed by competent authorities (Article 5): The choice of measure is up to each platform. Platforms will need to ensure that the measures are effective in tackling terrorist content without having adverse impact on human rights and freedom of speech.
    - Preservation. Companies are obliged to preserve removed terrorist content for six months (Article 6).
    - Transparency reporting. Companies will need to produce transparency reports on measures taken to comply with the regulation (Article 7). Platforms will also need to describe more widely the efforts they are making to remove terrorist content.
    - Complaint mechanisms. Platforms will need to introduce complaint mechanisms for users whose content has been removed (Article 10)
    - User notice (Article 11). Companies will need to inform users when their content has been removed as a part of the company complying with the regulation
    - Point of contact. Platforms will have to establish a point of contact to coordinate and respond to removal orders from competent authorities (Article 15).
    - Legal representative. All non-EU based platforms offering services in the EU will need to assign a legal representative in the EU (Article 17).
    - Violations of the above obligations may lead to penalties (Article 18), although the regulation does not specify what exact penalty would be awarded outside of violations of Article 3.

- Digital Services Act: proposed measures
  - Modification of the E-Commerce Directive's liability scheme, which sees platforms largely protected from liability but could leave them liable if they have "actual knowledge" of illegal content on their sites (Articles 3-5). There is also a provision which specifies that platforms will not be held liable for proactively carrying out activities aimed at reducing the presence of illegal content (Article 6).
  - Smaller platforms are largely exempt from some of the more rigorous requirements, whereas so-called "very large online platforms" (platforms with more than 45 million monthly active users in the EU) will need to introduce additional measures, including additional transparency obligations, risk assessments and being subject to independent audits (Articles 25-33).
  - Content removal order mechanism (Article 8).
  - Requirements to assign points of contact and legal representatives in the EU (Articles 10-11).
  - Obligations to clarify content moderation policies and practices in Terms of Service (Articles 12).
  - Transparency reporting obligations. Smaller and micro companies (as specified in Recommendation 2003/361/EC) are exempt in the proposal (Article 13).
  - Notice and action mechanism, allowing anyone to report suspected illegal content to platforms. Such a report would qualify as "actual knowledge" and would therefore render platforms liable for hosting such content (Article 14).
  - Trusted flaggers. Platforms will have to ensure that trusted flaggers' reports are prioritised (Article 19).
- Companies have the possibility to participate in several 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.

TECH AGAINST TERRORISM COMMENTARY

Regulation on addressing 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. The Commission's proposal drew criticism from <u>academics</u>, experts, and <u>civil society groups</u>. Further, the proposed regulation was criticised by <u>three separate UN Special Rapporteurs</u>, the <u>Council of Europe</u>, and the EU's own <u>Fundamental Rights Agency</u>, which said that the proposal is in possible violation of the <u>EU Charter for Fundamental Rights</u>. 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. 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 <u>welcomed</u> by civil society groups. In April 2021, the regulation was approved by parliament following Council review.

At Tech Against Terrorism, we have throughout the trilogue stage highlighted <u>our concerns</u> over smaller platform capacity and effectiveness in achieving its intended purpose of tackling terrorist content online and creating a safer EU. We have also shared several of the freedom of expression concerns that have been raised by civil society groups. At the publication of the final regulation in 2021, we <u>reiterated</u> these concerns, and highlighted that the regulation provides almost no legal certainty for platforms. It also offers little clarity on how smaller platforms will be supported in tackling this threat, and in complying with the regulation.

We believe that this is a misjudgement from the EU that casts further doubt over what evidence basis underpins the regulation. As a result, we fear that the regulation will do little to achieve its intended purpose of tackling terrorist use of the internet and risks harming innovation and competition in the process. Furthermore, the EU should clarify what safeguards are in place to avoid authorities abusing their position, and consider the incentives the law creates and what this means for the objectives the EU has set out in its overall tech strategy and the Digital Services Act.

The Digital Services Act

The Digital Services Act is an ambitious proposal aimed at introducing several new regimes to tackle illegal and harmful online content in the EU. It was relatively well-received by civil society groups, with groups commending the focus on transparency, accountability, size-oriented obligations, and that the liability scheme set out on the E-Commerce Directive remains largely intact, albeit slightly modified. However civil society groups criticised certain aspects of the proposal. For example, Electronic Frontier Foundation criticised the fact that notices in Article 14 equals actual knowledge, noting that this may lead companies to err on the side of removal, since such a notice can make them liable. This mechanism might force companies, as opposed to courts, to act as arbiters of legality. Article19 added that there are not sufficient human rights safeguards built into the risk assessment and audit provisions assigned to very large online platforms.

At Tech Against Terrorism, we noted that the DSA has several positive aspects, but it is unlikely that it will contribute to preventing terrorist use of the internet. In our assessment, the DSA is – whilst more balanced than other regulation – part of a global trend in which governments and inter-governmental bodies implement mechanisms that risk undermining the rule of law. Furthermore, despite claiming to want the opposite, the DSA may give private and democratically unaccountable tech platforms more power over online speech. The DSA is also part of another global trend in that it risks leading to increased extra-territorial enforcement of national law. There are risks that the DSA will lead to a more fragmented regulatory landscape in the EU, rather than harmonising it.

In our response to the draft DSA, we highlighted that governments and bodies like the EU should provide strategic leadership on matters related to terrorism, both online and offline. The DSA in our view, does not do that, and (whilst containing several commendable aspects) focusses on the wrong issues in terms of tackling terrorist use of the internet. Instead, the DS should focus on improving designation of far-right terrorist groups, supporting smaller tech companies (where most terrorist activity is located) in tackling terrorist use of their platforms, and formulate effective and human rights compliant approaches to tackle terrorist operated websites.

Annex 1. 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[73] 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.

---

73. Such as Council Framework Decision 2002/475/JHA

Annex 2. Online terrorist content: foundational regulation

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.[74] Further, Article 15 outlines that tech companies 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. This means that in theory, the AVMSD could apply to social media platforms on which videos are shared, including livestreaming functions.

---

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

Annex 3. EU-led voluntary collaborative forums to tackle terrorist use of the internet

- <u>EU Internet Forum (EUIF)</u>, bringing together Member States, tech companies, and relevant expert stakeholders[75] 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. Professor Danielle Citron described the EUIF as an example of the EU contributing to "<u>censorship creep</u>".[76] According to Citron, several of the voluntary steps that tech companies have taken to address terrorist use of their platforms 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, which negatively impacts 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.

---

75. Tech Against Terrorism has participated in EUIF meetings since 2017
76. By censorship creep, Citron means that online counterterrorism efforts or mechanisms risk taking on functions beyond its intended purpose, which risks leading to censorship of legal and legitimate speech online.

- <u>EU Internet Referral Unit</u> (<u>EU IRU</u>), based on the model pioneered by the UK's Counterterrorism Internet Referral Unit. The EU IRU employs subject matter experts to 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 <u>a joint operation with Telegram</u> to remove a large number of Islamic State channels. According to the EU IRU, the Unit has to date <u>referred more than 111,000 pieces of content</u> to tech companies. Whilst this approach has been commended, criticism has been leveraged against the EU IRU (and IRUs generally) as they risk undermining the rule of law. This is because they can promote 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 <u>noted</u> that there is no formal oversight of judicial review of the EU IRU's activities.

EUROPE | THE UNITED KINGDOM

Amongst the different global key trends identified by Tech Against Terrorism, the UK follows:

- Encouraging increased reliance on automated moderation tools
- Holding platform employees legally liable
- Mandating transparency and accountability

In 2019, the UK outlined an ambitious plan for online regulation in the Online Harms White Paper, aiming to make the UK "the safest place in the world to be online". The Paper aims to counter various online harms ranging from cyberbullying to terrorist content. The 2019 Paper was followed in 2020 by the adoption of The Interim Code of Practice on Terrorist Content and Activity Online and The Interim Approach for regulating video-sharing platforms, and in May 2021 the draft Online Safety Bill was published. The UK also has extensive counterterrorism legislation, including the criminalisation of viewing and sharing terrorist content online.

UK's regulatory framework

- The Draft Online Safety Bill (OSB) presented to Parliament by The Department for Digital, Culture, Media and Sport (DCMS) in May 2021. The Bill outlines duties imposed on user-to-user services and search services, to limit illegal and "harmful" content online. The Bill aims to safeguard journalistic and "democratic" content. The UK first set out its intentions for an Online Safety Bill in the Online Harms White Paper, published in May 2019.
- The Interim Code of Practice on Terrorist Content and Activity Online, December 2020, provides detailed guidance for tech companies to counter terrorist content and terrorist exploitation of their platforms until the Online Safety Bill is passed by Parliament.
- The Interim Approach for regulating video-sharing platforms (VSPs), November 2020, regulates VSPs services until the OSB is passed, and came into effect on 1 November 2020. It has transposed the VSP framework into Part 4B of the Communications Act 2003 ("the Act").

- <u>The Terrorism Act 2000</u> is a cornerstone of UK terrorism legislation. Section 58 of the Act specifies the offence of possessing information, including via online means, that is <u>"useful to a terrorist"</u>.
- <u>The Terrorism Act 2006</u> creates new offences related to terrorism, as well as amends existing ones. A relevant example is <u>Section 2</u>, which makes it an offence to disseminate terrorist propaganda for "terrorist purposes".
- <u>The Counter-terrorism and Border Security Act 2019</u> amends section 58 of the Terrorism Act. It also criminalises obtaining or viewing such material online.

Main body overseeing online regulation

- <u>Ofcom</u>, the UK communications regulator. Ofcom oversees the application of new regulations related to online platforms, both under the Interim Approach and the draft Online Safety Bill.

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.
- <u>The Department for Digital, Culture, Media and Sport (DCMS)</u> 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 and presented the Draft Online Safety Bill to parliament.
- The <u>Home Office</u> is responsible for security and policing in the UK, including counterterrorism and terrorist use of the internet.
- <u>The Independent Reviewer of Terrorism Legislation</u> scrutinises and reports on terrorism legislation in the UK. The current reviewer is Jonathan Hall.

Key takeaways for tech companies

Draft Online Safety Bill

- The Bill imposes duties of care on online platforms to moderate user-generated content. This applies to both user-to-user services and search providers.
- The Bill will require platforms to:
  - Mitigate and manage risk of harm that users may experience due to illegal content on their platform.[77] Companies need to carry out the following duties:
    - Use proportionate systems designed to minimise the presence of illegal content, minimise the time in prioritising illegal content is on the platform, and minimise the dissemination of priority illegal content.
    - "Swiftly" remove illegal content that was alerted by another party. The Bill does not define a specific timeframe for what counts as "swiftly" removing content.
  - Specify how users are protected from illegal content in their Terms of Service (ToS) and a duty to lay them out clearly and apply them.
  - Conduct risk assessments to assess the presence of illegal content on their platform, including a new risk assessment every time a platform changes its policies or operations that could affect the illegal content on its platforms. These requirements will differ somewhat between user-to-user services and search providers. When user-to-user services and search providers conduct their risk assessments, a few criteria need to be considered.[78]
  - Pay due regard to freedom of expression, privacy, and to protect "democratic" and journalistic content.[79]

---

77. Illegal content is defined as (a) in relation to a regulated user-to-user service, content— (i) that is regulated content in relation to that service, and (ii) that amounts to a relevant offence. It also includes terrorism content which is defined as "any offence under the provisions of the Terrorism Act 2000, section 113 of the Anti-terrorism, Crime and Security Act 2001 and the Terrorism Act 2006".

78. User base, Levels of risk that users face from particular harms or types of content, including such as: terrorist content, child sexual exploitation and abuse content, "priority illegal content, other illegal content – the Bill does not specify what kind. What algorithm the platform uses and how fast and wide content spreads across the platform. The Bill does not specify how platforms should do so.

79. This is done through the following duties: Duty to protect users' right to freedom of expression within the law; Duty to protect users from unwarranted infringements of privacy when deciding on, and implementing, safety policies and procedures; Duty to include impact assessments of safety policies and procedures on freedom of expression and privacy; Duty to publish how platforms protect users' freedom of expression and right to privacy in a platform's ToS.

- The Bill also obligates tech platforms to publish annual transparency reports. These must include information on the notices that Ofcom serve to a tech platform, information on illegal content on the platform, information on redress mechanisms, information on the safety measures taken by the platform.[80]
- The Bill imposes liability on tech platforms and tech platform employees when platforms do not fulfil the obligations of the Bill.[81]
- The Bill specifies obligations that Ofcom should provide guidance on how tech companies should uphold all duties. In addition, the Bill notes that when assessing tech companies' ability to fulfil such obligations, size and capacity of a platform will be taken into account.[82]

---

80. Notices come in many forms in the Bill, such as technology notices and penalty notices. They are served by Ofcom to tech companies and demand information on a particular topic such as illegal content or the use of technological processes or features.

81. More requirements are specified in the Bill, please get in touch if you are a tech platform and would like further guidance on transparency reporting and UK legislation.

82. This liability comes in three forms: senior managers liability, individuals liability, and parent company liability.

Interim Regime for Regulating Video-Sharing Platforms (VSPs)

- Ofcom expects VSPs to assess whether they fall in the remit of the new legislation, and to conduct risk assessments to identify what the potential harms are to their users:
- "Relevant harmful material" in the Interim Regime is defined as "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" and "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". It also captures content that can be seen as incitement to violence and hatred.
- VSPs, regardless of their size, need to protect users under the age of 18 from accessing restricted material.[83] Platforms need to regulate such content based on "proportionality", and need to consider the size and nature of the service, the type of harm caused, the exposed user's characteristics that might be protected (such users who identify as LGBTQ+), and the implications for freedom of expression. Platform size is assessed by company user base.
- VSPs need to implement a user appeal mechanism for users whose content has been removed.
- VSPs need to implement an independent redress mechanism, which will give users the ability to submit a final appeal to an independent body to review the platform's initial decision.
- 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". However, 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. It is unclear what will happen when the "early regulatory period" ends.
- Tech Against Terrorism offered a response to Ofcom's consultation process on the regulation of VSPs, which was concluded in September 2020, which can be found here.
- Ofcom has opened a 2021 consultation and welcomes responses from stakeholders. The consultation can be found here.

---

83. 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".

Interim Code of Practice on Terrorist Content and Activity Online:

- The Interim Code is grounded in five principles that the UK government believes tech companies should comply with. The Code applies to companies that host user-generated content on their website.[84] Tech companies are expected to:
    - Identify and prevent the dissemination of terrorist content and use of tech companies' platforms in the UK to protect UK citizens from accessing such material.
    - Minimise the potential search results linking to terrorist content and activity.
    - Take part in cross-industry collaborations to find effective and coherent solutions to counter terrorist use of the internet.
    - Implement user reporting, complaints, and redress mechanisms to ensure users are empowered and their rights are protected.
    - Support the UK authorities in investigating and prosecuting terrorist offences made by individuals under existing Terrorism Acts.
- The Interim Code defines terrorist content as: "any content which, by uploading it or otherwise making it available to others online, a person is committing an offence under UK terrorism laws.[85] It also includes content produced by terrorist organisations proscribed terrorist in the UK.
- The Code states that tech companies should not end their efforts to counter terrorist use of the internet "if a company addresses content and activity on the basis that it suspects on the balance of probabilities the content and activity to be terrorist in nature, they will not have to prove beyond a reasonable doubt that the content and activity would constitute that offence."
- The Code includes guidance for "small and medium size enterprises" (SMEs), as it specifies that SMEs need to assess themselves which of the five principles of the Code applies to them, and which recommended measures may prove effective in countering terrorist exploitation on their platforms. The Code also sets out the minimum harms that an SME should consider, which are:
    - Targeted radicalisation of vulnerable users.
    - Sharing of terrorist content (including propaganda through all media types).
    - Posting of URLs to terrorist content and third-party services.
    - Live broadcast of terrorist activity.

---

84. The Code defines user-generated content as "digital content that is produced, promoted, generated or shared by users of a service that is managed and/or owned by a third party. The Interim Code of Practice on Terrorist Content and Activity Online, p. 16.
85. The Interim Code of Practice on Terrorist Content and Activity Online, p. 16

- The legislation highlights existing efforts to counter terrorist use of the internet and the dissemination of terrorist content, including the work by Tech Against Terrorism, the Global Internet Forum to Counter Terrorism, the Christchurch Call to Action, and TechUK.

TECH AGAINST TERRORISM COMMENTARY

Circular definitions of illegal and terrorist content

Tech Against Terrorism cautions that the OSB's definitions for both illegal content and terrorist content are impractically broad and circular. Illegal content is in the draft Bill defined as content that leads to an offence, and terrorism content is defined as "terrorism content that leads to a terrorist offence". This definition does little to inform a tech company about what content falls under these definitions, nor does it inform how tech companies should operationalise this definition when acting against terrorist exploitation of their services.

This leaves tech companies to adjudicate on what constitutes terrorist content. Whilst terrorist content that clearly depicts violence and incites violence might be easy for platforms to detect, in reality most terrorist groups frequently share more "grey area" content, which is generally difficult to identify. The current definition means that tech companies will need to make difficult decisions in correctly assessing whether content is terrorist or not. Beyond this, the Interim Code specifies that tech companies are expected to go beyond this definition, and act when there is "reasonable doubt". We deem that such an expectation is too vague for tech companies to comply with. This approach outsources the responsibility of adjudication of what is acceptable speech to tech companies.

We understand that Ofcom will provide further guidance to tech companies on how they need to uphold the Bill's obligations on illegal and terrorist content. However, we still deem that the law itself should be more narrowly defined to not put the full weight of implementation on Ofcom and tech companies. For these regulations to target what they set out to, we advise more clear and detailed definitions.

"Democratic" and journalistic content: lack of definitional clarity

Tech Against Terrorism welcomes the fact that the OSB recognises that in moderating online speech, journalistic and otherwise legitimate material may be erroneously removed. This has negative consequences for human rights and freedom of speech. However, we caution the way in which the UK government endeavours to protect such material online.

The Bill specifies "democratic content" as any content that furthers democratic debate in the United Kingdom. Due to this vague definition, Tech Against Terrorism is concerned that tech companies will struggle to adjudicate on what is illegal or terrorist content as opposed to democratic. Without clear definitions and detailed guidance, individual tech companies have to interpret this, which in turn risks rendering inconsistent application of the law.

Likewise, journalistic content is defined in the Bill as content that may be considered journalistic in nature. Tech companies are likely to struggle to adjudicate on what is terrorist content that is shared for terrorist purposes versus terrorist content that is shared for journalistic objectives. The UK government should provide more context and guidance to tech platforms on how to determine what is journalistic content.

Vague definitions risk being weaponised by terrorists and violent extremists

To avoid content moderation by tech platforms, terrorists and violent extremists use a variety of content moderation avoidance techniques. We worry that the broad definitions of journalistic and democratic content can be exploited for this purpose. In our research we note that both Islamist and far-right terrorist and violent extremists avoid content moderation by posing as news agencies or journalists. This presents challenges to tech companies to identify whether content is journalistic or terrorist. Further, violent extremists or terrorists may deem that their online content furthers the democratic debate (see definition of "democratic content above) and this could provide a justification for them to appeal their content being taken offline.

Further detail and guidance on how to differentiate between legitimate journalistic and democratic content and terrorist content is required to support platforms in meeting the required duties effectively.

Outsourcing the responsibility to adjudicate to tech companies

Due to the broad, circular, and oftentimes vague definitions of illegal, terrorist, harmful, democratic, and journalistic content, as well as the absence of judicial oversight in content moderation, the responsibility of adjudicating is left to tech companies. This has resulted in a process in which democratically unaccountable tech companies have set online speech standards. In our view, it is vital that counterterrorism, whether offline or online, is led by democratically accountable institutions in accordance with the rule of law. The above definitions should be clearer to inform tech companies on what content should be taken offline and what should remain.

Broad definitions and enforcement mechanisms risk freedom of speech

The broad definition of online harms and terrorist content will leave tech companies with the responsibility of adjudicating what constitutes illegal and harmful content without guidelines on how to assess such material. Due to the high fines and legal liability faced by tech companies if they fail to identify and address content effectively , 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 content, which severely undermines the rule of law and freedom of expression. In our view, speech that is legal offline should not be criminalised online. If the UK seeks to criminalise speech online, it should ensure that the offences have a basis in criminal legislation.

Tech company liability

The OSB proposes legal liability for managers, employees, and parent companies if they cannot show the reasonable steps they took to remove illegal and terrorism content of the internet. This puts the responsibility of terrorist content on tech companies when this should be attributed to the terrorists and violent extremists that disseminate such material online.

The draft OSB specifies that even if tech companies take measures against terrorist content, Ofcom will still consider whether terrorist exploitation is still "prevalent" on a particular platform. This implies that if terrorist content is still prevalent, it will not be enough for Ofcom to consider that tech companies fulfilled their safety duties. We deem that if a tech company acts and attempts to moderate terrorist content of their platform, this should be sufficient in guaranteeing that they are seen as upholding their safety duties. Terrorists and violent extremists are known to adapt their content dissemination strategies or platforms of choice to circumvent content moderation, and it would be unfair to punish a platform for this.

To preserve the freedoms that the internet enables, and the competition that drives innovation and new forms of expression and communication, we must be cautious of the risks that misguided modification of current legal liability schemes may pose.

The OSB poses risks smaller tech companies and an open and vibrant internet.

Whilst we welcome that the Online Safety Bill and the Interim Code mentions smaller tech companies and their lack of resources when outlining the minimum set of online harms smaller companies need to monitor, it still sets out ambitious requirements for smaller platforms. For example, the requirements laid out in the Interim Code will be complicated for tech companies to fulfil. An example would be for a small tech company to stay abreast of URLs containing terrorist content that are posted on their platform by third-party services, especially for a small platform with limited resources. In addition, "targeted radicalisation of vulnerable users" will also be hard to assess for tech companies in any case, therefore especially for smaller platforms.

Policymakers should consider platform size to ensure small companies are not disproportionality affected by regulation

Encouraging risk assessments: a welcome step provided support is available
Tech Against Terrorism carefully welcomes the UK government's focus on risk assessments and deem this to be the first step in countering terrorist use of the internet. Tech companies should to the best of their abilities consider how their platforms could be exploited and remain aware of potential adversarial shifts. Tech companies, especially smaller platforms, will need support in carrying out risk assessments, as they may not have the resources or capacity to conduct these. It is our hope that the UK government will support smaller tech companies in responding with terrorist exploitation of the internet.

Resources

Article 19, (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 Natasha (2019), UK Sets Out Safety-focused Plan to Regulate Internet Firms, Techcrunch

Osborne Clarke (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.

UK government (2020), Online Harms White Paper - Initial Consultation Response

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

EUROPE | TURKEY

Amongst the different global key trends identified by Tech Against Terrorism, Turkey follows:

- Mandating different requirements depending on platform size
- Mandating short removal deadlines
- Mandating a local presence

Online content regulation in Turkey is characterised by extensive removal of material that has resulted in the government blocking a large number of Turkish and international websites. Furthermore, the Turkish government recently introduced the Social Media Bill in October 2020, which implements a wide range of new requirements and steep penalties for social media companies, which threatens freedom of expression in the country, according to critics.

Regulatory framework

- Many provisions of the Criminal Code and other laws, such as Turkey's Anti-Terrorism Law and Defamation Law are applied 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. These include storing user data in Turkey, shorter timeframes for responding to complaints about posts that violate personal and privacy rights, as well as fines for failure to comply.
- 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 online content, such as child abuse images and obscenity, and enables the blocking of websites.

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 defence, terrorism, international crimes, and cybersecurity passing through telecommunications channels", without a requirement to obtain a court order.
- The Information and Communication Technologies Authority (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 must obtain an "activity certificate" to legally operate in Turkey, while those who are not members are not able to provide services within the country.
  - Blocking orders can be issued by courts, public prosecutors, or the BTK.
    - Websites hosted in Turkey that contain 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 of containing 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, founder and first president of the Republic of Turkey.

- 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 of up to around $52,150.[86] If ISPs fail to block all alternative means of accessing the targeted site, such as proxy sites, it could result in a fine of up to around $8,690.[87]
- 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 around $700,000.
  - International companies are required to store user data in Turkey.
  - The Bill allows courts to order Turkish news websites to remove content within 24 hours.
  - If social media companies do not comply with all 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 steep 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 around $17,380[88] 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. The Law states that:
  - Public and private bodies, such as banks, archives, professional organisations, and private companies, must when requested by the MIT data, documents, or information pertaining to 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. Access providers are required to provide assistance to the TIB (since 2016, the BTK) in monitoring internet traffic.

---

86. i.e. 300,000 Turkish liras
87. i.e. 50,000 Turkish liras
88. i.e. 100,00 Turkish liras

TECH AGAINST TERRORISM COMMENTARY

Broad or vague definitions & extensive blocking powers

Turkey's Anti-Terrorism 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.[89] Tech Against Terrorism is concerned that imprecise definitions of terrorism could encourage tech platforms to remove content that is shared with the purpose of documenting terrorist offences and war crimes, which can serve as crucial evidence in court proceedings. Additionally, governments could take advantage of imprecise definitions on terrorism in order to censor their citizens – ultimately infringing upon freedom of expression.

User privacy risks

The new Social Media Bill introduces a strict regulatory framework for tech companies operating in Turkey. Under the new 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 within 48 hours or receive steep fines, and international companies are required to store user data in Turkey.

Whilst it is good that these requirements do not apply to small platforms, mandating user data to be managed within the jurisdictions of the country presents risks over user privacy. This is because the government and law enforcement would no longer have to go through Mutual Legal Assistance Treaties (MLAT) to request platforms to disclose information about their users.

90

---

89. 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.

Steep penalties

If social media companies do not comply with the new criteria within six months of the legislation going 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%.[90] 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. Further we worry that punishments such as banning advertising on the platforms or adjusting the bandwidth, in addition to high fines, will severely harm competition and innovation. While it is the major tech companies who are meant to face legal sanctions and service interruptions under Turkey's Social Media Law, the country's citizens risk also suffering consequences.[91] Examples of the law's potential repercussions on citizens include slower Internet services, navigated cowed social platforms, and authorities targeting dissent or anti-government critics.

---

90. Civil society has also voiced concerns about the new rules, which impose heavy fines, ad bans, and tighten bandwidth if social media companies fail to meet the 6-month deadline to adhere to the new criteria, which could expand government surveillance of citizens as well as make companies complicit in undermining civil liberties.
91. Read more on the EFF's argument here.

SECTION 3 |
GLOBAL ONLINE REGULATION

NORTH AMERICA

NORTH AMERICA | CANADA

Canada's approach to online regulation has, so far, supported 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 about introducing legislation on harmful online content, and potentially making tech companies liable for content on their platforms.

Canada's regulatory framework

- Canada is a signatory to the Christchurch Call to Action.
- Canada's Communications Future: Time to Act (BTLR), January 2020, is 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.
- National Strategy on Countering Radicalization to Violence, 2018, summarises Canada's approach to countering terrorism and violent extremism.
- Canada's Digital Charter, 2019, 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.
- In January 2021, Heritage Minister Steven Guilbeault announced that a new regulatory framework for tech platforms is expected be introduced in 2021 in the House of Commons. This intention was set out in the Mandate Letter from Prime Minister Justin Trudeau, which tasked Minister Guilbeaut with developing 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."

- In June 2021, the government of Canada <u>announced</u> new regulatory measures to "better protect Canadians from hate speech and online harms". The initiative – introduced by the Department of Justice, the Department of Canadian Heritage, and Public Safety Canada – aims to tackle "the most extreme and harmful speech" both online and offline. The proposed legislation will amend the Canadian Human Rights Act, the Criminal Code, and the Youth Criminal Justice Act to redefine hate speech and hatred. These amendments will also provide additional tools to prevent and offer remedies to hate speech and hate crimes. A definition of "hatred" will be added to section 319 of the Criminal Code. In addition, the government of Canada will introduce legislation to tackle harmful content online. This legislation will cover terrorist and hate speech content, as well as content inciting to violence.
- <u>Bill C-10 amending the Broadcasting Act</u>, 2021: In 2019, policymakers began discussing Bill C-10 to amend the existing Broadcasting Act, and allow for the Canadian Radio-Television and Telecommunications Commission (CRTC) to regulate online streaming services and promote Canadian content. At the time of writing, Bill C-10 is still being discussed in Canada, and whether it will cover user-generated content and how remains unclear.[92]
- Bill C-10 is the first of a set of legislative frameworks meant at increasing government-led regulation of online platforms and content in the country, paving the way for the possibility to mandate content moderation policies and process.

---

92. Originally, social media content was to be excluded from Bill C-10, which focused on broadcasting and streaming platforms. However, the first quarter of 2021 was marked by heightened discussion in Canada regarding Bill C-10's scope of application, with the Heritage committee's (in charge of the drafting bill) decision to remove a clause exempting user-generated video content from the Bill. Following this, an amendment was added with the current version of the Bill, as of May 2021, stating that Bill's scope of applicability over social media content is limited to promote the discoverability of content by Canadian Creators.
See: Karadeglija Anja (2021a), <u>Bill C-10 amendment that would exempt social media content from regulation voted down</u>, The National Post.

Relevant national bodies

- The 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 and safety, including counterterrorism. Public Safety Canada also runs the 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 and the drafting of the upcoming regulation for online platforms.

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.
- However, this is expected to change with a new regulation[93] likely to be proposed by Canadian Heritage. It is unclear what this regulatory framework will introduce, though it will likely be modelled after Germany's NetzDG. Currently, no draft regulation has been presented, and no public consultation has been held. However, different civil society organisations have published reports papers outlining what a regulatory framework could look like. A few key facts are currently known:[94]

    - Illegal content is likely to be divided in 5 categories, including terrorist content, hate speech, and content that incites violence.
    - A new definition of hate is likely to be laid out based on case law, with new rules to counter hate speech.
    - Tech companies are likely to be required to monitor and remove illegal content on their services. Deadline for removal is likely to be 24 hours.
    - Companies would be fined for non-compliance, and a new regulator will be set up to oversee implementation and compliance.
    - A new regulator would be created to guide tech companies on the new definition of hate and the appropriate actions to take.
    - 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.[95]

---

93. See: Silver Janet E. (2021) Regulation of Online Hate Speech Coming Soon, Says Minister, iPolitics.ca; Elghawaby Amira (2021), Canada is Bringing in New Legislation to Stop the Spread of Online Hate. Here's how it can work; and Leavitt Kierann (2021), After the Capitol riots, Ottawa draws lessons about social media regulation, Toronto Star.
94. See: Elghawaby Amira (07.04.2021), Canada is Bringing in New Legislation to Stop the Spread of Online Hate. Here's how it can work; and Karadeglija Anja (2021b), New definition of hate to be included in Liberal bill that might also revive contentious hate speech law, The National Post.
95. At the time of writing, there are still uncertainties about whether the recommendations made in the BTLR are to become laws in Canada.

TECH AGAINST TERRORISM COMMENTARY

A shifting approach to online regulation

Canada's introduction of a formal framework that mandates tech companies to monitor and remove certain illegal and harmful content represents a significant shift in the government's approach to online regulation, which had favoured self-regulation and cross-sector initiatives. This shift reflects the recommendations laid out in the BTLR, which called for a "legislation with respect to liability of digital providers for harmful content and conduct using digital technologies" to counter the spread and amplification of "harmful content".[96] Proposed changes to the online regulatory framework have been criticised by digital right experts and by the country's political opposition.[97] Criticisms have focused on the risks for freedom of expression both with regard to requiring tech companies to rapidly adjudicate on the legality of content and promptly remove it within 24-hours, and with the requirements that platforms promote "Canadian" content.

At the time of writing,[98] it is too early to assess what the upcoming legislations being discussed in Canada will mean in practice for content regulation and tech companies. It is also currently unclear whether the legislations will provide effective support for tech companies in disrupting the spread of terrorist content. Canada has a track record of providing practical support to the tech sector in countering terrorist use of the internet, notably via Public Safety Canada's funding of the Terrorist Content Analytics Platform. Tech Against Terrorism recommends Canada to continue supporting practical and policy assistance to tech companies in countering terrorism whilst respecting human rights, and to include this support in future regulations on online content. We also caution Canada to be careful in replicating legislations passed elsewhere, and to consider the criticisms raised by counterterrorism experts and digital rights advocates regarding these legislations.

---

96. A term undefined in the BTLR, the scope of harmful content to be addressed by tech companies, appears to be limited to five categories of illegal content: "Hate speech, terrorist content, content that incites violence, child sexual exploitative content and non-consensual sharing of intimate content."
Elghawaby Amira (2021), Canada is Bringing in New Legislation to Stop the Spread of Online Hate. Here's how it can work., Press Progress.
97. Changes to the regulatory framework are proposed by the current Liberal government of Prime Minister Justin Trudeau, and have been opposed by the country's political opposition parties, notably the Conservatives.
98. April 2021

NORTH AMERICA |
THE UNITED STATES
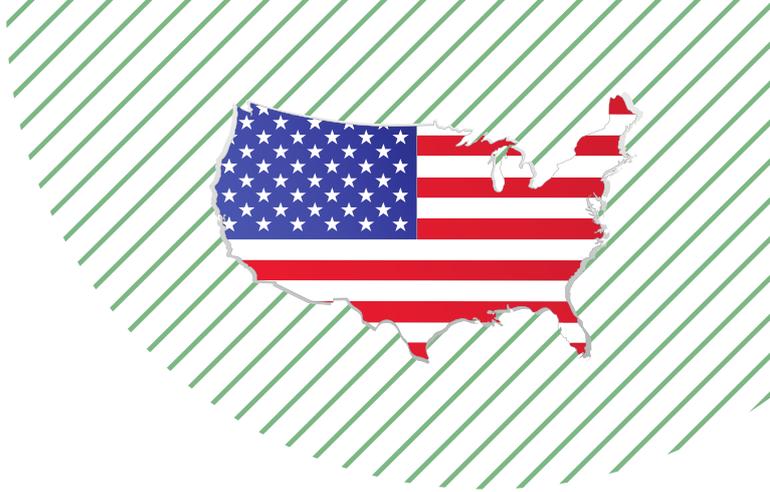
Online regulation and content moderation in the United States is defined by the First Amendment right to freedom of speech and Section 230 of the Communication Decency Act 1996, which establishes immunity from legal liability for tech platforms. It has broadly impacted the innovation of the modern internet and has created lasting effects beyond the US. The former Trump Administration administered an executive order in May 2020, that directed 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 resulted in a wave of proposed bills and Section 230 amendments from both government and civil society.

US regulatory framework

- <u>First Amendment</u> 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.
- <u>Section 230 of the Communication Decency Act of 1996</u>, 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:

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

Key takeaways for tech companies

- First Amendment law 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. This 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.[99]
- However, this might change in the coming years. Since 2018, there have been several proposals made from members of Congress, and from former President Donald Trump, to alter Section 230. There is currently a bipartisan bill that has been reintroduced in early 2021, for Section 230 which experts say have a chance of passing:
- The Platform Accountability and Consumer Transparency Act (PACT Act), originally introduced in June 2020, focuses on promoting platform transparency and accountability. An updated version was reintroduced in the 2021-2022 congressional session with changes such as carving out expectations for individual providers, outlining scalable requirements based on a platform's revenue and size, and clarifying platform obligations regarding the complaint system, the phone line, and transparency reporting.[100]
- To access a comprehensive list of the proposed legislation so far, we recommend the Section 230 Reform Legislative Tracker, which includes information on each bill that has been introduced in Congress to reform Section 230 since 2020.

---

99. 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.
100. Future Tense (2021), All the Ways Congress Wants to Change Section 230, Slate.

TECH AGAINST TERRORISM COMMENTARY

First Amendment and Section 230 "Impunity"

There is an active discussion about Section 230 and its shortcomings, however, "most experts agree that it has been the bedrock of the growth of the internet sector since it became law in 1996 and that it is a cornerstone of online expression", by protecting platforms from certain types of liability for user-generated content.[101]

Policymakers across the political spectrum have criticised Section 230. On the political left, some argue it has enabled tech platforms to host harmful content with impunity, while on the right, some argue that it has enabled tech platforms to disproportionately suppress conservative speech. Both sides agree that Sections 230 needs to be updated and reformed.[102] While some of the reform proposals such as the PACT Act (see below) are reasonable according to experts and critics, others collide with the First Amendment.[103]

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. For example, one proposal comes from Danielle K. Citron, a law professor at Boston University. Citron has suggested amending 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". A judge would assess the "reasonableness" of a platform's overall policies and practice at a preliminary stage of a lawsuit.

---

101. Future Tense (2021), All the Ways Congress Wants to Change Section 230, Slate.
102. ibid.
103. https://www.eff.org/deeplinks/2020/12/its-not-section-230-president-trump-hates-its-first-amendment

PACT Act: Rule of Law Implications

The PACT Act is one of many proposed reforms to Section 230. Experts and critics have claimed the PACT Act seems to be the most reasonable proposal and likely to be passed.[104] The PACT Act – which was underlined updated in March 2021– contains "the same fundamental flaws as the original: creating a legal regime that rewards platforms for over-censoring users'speech".[105] The PACT Act therefore, does not require takedown notices to be based on final court order or adjudications that have found content to be unlawful or unprotected by the First Amendment.[106] Given that final orders on the issue of legal speech issued by lower courts are often reversed by appellate courts, the PACT Act could risk the takedown of lawful content.[107] Tech Against Terrorism cautions that the Act, if introduced, should avoid introducing measures that risk undermining the rule of law, for example by removing legal content or contributing to extra-legal norm-setting.

104. Harmon Elliot, (2020), It's Not Section 230 President Trump Hates, It's the First Amendment, EFF.
105. Mackey Aaron (2021), Even with Changes, the Revised PACT Act Will Lead to More Online Censorship, EFF.
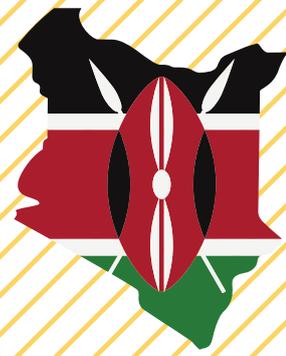106. ibid.
107. ibid.

SECTION 3 |
GLOBAL ONLINE REGULATION

MENA &
SUB-SAHARAN AFRICA

MENA & SUB-SAHARAN AFRICA | KENYA

Amongst the different global key trends identified by Tech Against Terrorism, Kenya follows:

- Mandating a local presence

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 government has increased its counterterrorism efforts, as well as online content regulation. Civil society groups have criticised many of Kenya's legislations for being too broad, vague, and for causing 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.

Kenya's Regulatory framework

- Kenya Information and Communications Act, (KICA), October 1998, is 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", if passed, will introduce stringent regulations on the use of social media in Kenya.
- The Computer Misuse and Cybercrimes Act, 2018, 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, 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 the government 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 enables national security bodies to intercept communications "for the purposes of detecting, deterring, and disrupting terrorism". The act also includes provisions on radicalisation as well as on the "publication of offending material".
- Guidelines implemented by the CA are set up to curb online abuse:
  - The guidelines prohibit political messages that "contain offensive, abusive, insulting, misleading, confusing, obscene, or profane language".
  - 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. It has also been used by service providers and other state agencies, such as the National Cohesion and Integration Commission (NCIC), to monitor 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 introduces obligations for social media users. The regulations include new requirements for the operators of social media platforms to obtain licenses and establish a physical office in the country. It further aims to place regulations on content published by social media users. The Bill was tabled in parliament in Oct 2019, but has not progressed at the time of writing. Tech companies should monitor for any developments on the Bill.

TECH AGAINST TERRORISM COMMENTARY

Broad or vague wording and definitions in legislation

Many of Kenya's legislations have been criticised by civil society for their "broadness", "vagueness", and potential "detrimental implications for freedom of expression".[108] The Prevention of Terrorism Act, for example, includes a provision on radicalisation which 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 such an offense would be subject to up to 30 years in prison. The overly broad terms are also present in legislation related to online content regulation, namely, when the CA implemented new guidelines, to curb online abuse. The guidelines prohibit political messages that "contain offensive, abusive, insulting, misleading, confusing, obscene, or profane language". These guidelines have been criticised for being too broad, which could then be used to limit online expressions.

Tech Against Terrorism is concerned that imprecise definitions of terrorism could encourage tech platforms to remove content that is shared with the purpose of documenting terrorist offences and war crimes, which can serve as crucial evidence in court proceedings. Governments may also seek to take advantage of imprecise definitions in order to censor their citizens – ultimately infringing upon freedom of expression online in Kenya.

Strictly regulated user-generated content

The "Social Media Bill" if enacted, could largely impact social media companies and their users in Kenya through strict regulations on user-generated content. For example, 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.

Strict regulations on user-generated content on social media platforms coupled with vague definitions and therefore violations risks making users overly-cautious in their speech, potentially leading to self-censorship. Furthermore, high fines and vaguely defined grounds might lead to companies being overly-cautious with their content removal or blocking, which could lead to heightened censorship and infringe upon freedom of expression.

---

108. https://freedomhouse.org/sites/default/files/2020-02/Final_PolicyBriefKenya_11_14_18.pdf

MENA & SUB-SAHARAN AFRICA | MOROCCO
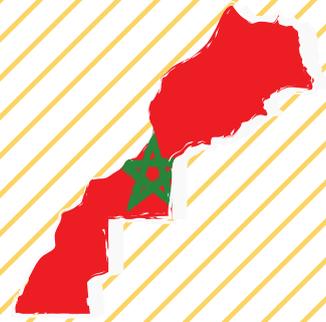
Morocco's online regulatory framework consists of different laws and codes that aim to limit the spread of content that poses a threat to the Kingdom's "integrity, security and public order". Central to this framework are the 2003 Anti-Terrorism Law, passed in the aftermath of the 2003 Casablanca bombings, and the 2016 Press Code that lays out limitations to journalistic publications and public speech. However, the existing regulatory framework is not explicitly clear on the 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

- Morocco's legal framework for countering terrorism,[109] May 2003, provides definitions of key terms (such as "terrorist acts") and laying out the different sanctions and legal processes. The law:
  - Prohibits and sanctions the diffusion of terrorist content by any means of speech (oral or written), including audio-visual and electronic material.
  - Prohibits and sanctions incitement to and condoning or legitimisation 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.

- Morocco's Press Code,[110] August 2016, regulates the press and public speech in general, including speech and journalistic content posted online. The Press Code specifies limitation to freedom of the press and public speech, and violation is subject to fines. Specifically, it prohibits the publication of content that threatens "public order", including anything that insults "Islamic religion, the monarchy, or the integrity of the Kingdom". Further, Article 72 penalises the diffusion, by any means (including electronic) and by all individuals of:

---

109. Loi n° 03-03 relative à la lutte contre le terrorisme.
110. Loi relative a la presse et a l'édition.

- Intentionally spreading allegations, and of false or falsified information, that have led to disruption to the public order or fear amongst population.
- Terrorism apology.
- Incitement to hatred and racial hatred.

- <u>Morocco's Penal code, 2018 consolidated version</u>, sanctions certain types of speech, including speech that is "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.
- <u>Draft law no. 22.20</u>, the so-called "social media law", passed in March 2020 and later temporarily suspended in May 2020 due to the ongoing Covid-19 pandemic. The draft law <u>would task "network providers"</u> with restricting access to and suppressing online content that poses a threat to security and public order within 24 hours.

Key takeaways for tech companies

- Under the current legislative framework, internet platforms are not liable for any 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.

TECH AGAINST TERRORISM COMMENTARY

Lack of a clear framework for content regulation

The 2003 counterterrorism law provides a legal basis for countering the dissemination of terrorist content by any means. However, the current counterterrorism and online regulation frameworks are unclear regarding the implications for tech platforms, and do not address whether platforms can be held responsible for user-generated content, or if they are protected by a safe-harbour provision. 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). Further clarification regarding government's regulatory power, and what is legally required of tech companies is warranted to strengthen due process surrounding online content regulation in the country.

MENA & 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. Jordan places the liability on internet users, rather than on tech companies. Some have criticised Jordan for unjustly applying laws that are aimed at combatting terrorism, hate speech, and disinformation, toward some forms of legally accepted speech, such as criticism of the government.

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, and widens 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.[111]
  - In May 2021, Minister of Parliament, Omar al-Naber, drafted a parliamentary memorandum that called on the government to include stipulations on "hate speech" on social media under Jordan's cybercrime laws.
- 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.

---

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

- The Telecommunication Act, 1995 regulates all telecommunication companies in Jordan and establishes the regulator. The Act criminalises the "illegal" use of public or private telecommunications networks, as stipulated by Jordanian penal codes.

Relevant national bodies

- The Telecommunications Regulatory Commission is responsible for regulating telecommunications and information technologies. It sets the policies operators need to comply with. It also grants licenses. As part of their mandate within the Cybercrime Law 2019, the Commission also oversees "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 Jordanian penal codes.
- The Ministry of Information and Communications Technology (MoICT) sets the policy directions for telecommunications and information technologies through biannual national strategic plans. The body coordinates with relevant stakeholders, and also 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 by the Cybercrime Laws.
- The General Intelligence Department (GID) – the country's intelligence agency which is in charge of national security. They are also involved with detaining suspected individuals, question and the monitoring of suspects, both offline and online.

Key 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

TECH AGAINST TERRORISM COMMENTARY

Broad definition of terrorism and potential human rights abuses

Civil society groups have raised concerns around the expansion of definitions of terrorism in Jordan's legislative frameworks.[112] More clarity should be given to users on what classifies as terrorist content. This is to ensure that the country's counterterrorism framework and legislation on hate speech cannot be used to restrict legal and non-violent speech.[113] This will ensure that Jordan's legislation effectively tackles terrorist and extremist use of the internet whilst upholding human rights, and particularly freedom of speech.

Jordan's enforcement mechanism and its potential infringement of the right of privacy

We have concerns over Jordan's enforcement mechanism, as the details of anyone suspected of terrorist activity online can be requested from an "application" or a public Internet café by the police and law enforcement agencies. In terms of "applications", 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 without a court order.[114] Users'right to privacy should be central when designing legislation to counter terrorist or extremist use of the internet.

---

112. <u>Human Rights Watch</u> pointed out how the definition can be used to quell not just expressions of terrorism, but also peaceful and legal speech. In addition, <u>Open Democracy</u> criticised the law for having been used in prosecutions of human rights activists and journalists.
113. With regards to the definition of hate speech, <u>AccessNow</u> mirrors the concern raised above as they 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.
114.   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.

SECTION 3 |
GLOBAL ONLINE REGULATION

LATIN AMAERICA

LATIN AMERICA | BRAZIL

Amongst the different global key trends identified by Tech Against Terrorism, Brazil would follow (if the Brazilian Internet Freedom, Responsibility and Transparency Act is passed):

- Outsourcing legal adjudication to tech companies
- Mandating a local presence

Brazil is the leading country in terms of internet and social media use in Latin America, and a major market for large tech platforms, including WhatsApp and Facebook. Brazil's approach to online regulation has been shaped by the purported disinformation campaigns that were coordinated on WhatsApp. The messaging app has been accused of being used "for the dissemination of fake news", and critics of the country's so-called "fake news" bill have said that WhatsApp served as a "standard" for new regulation of messaging apps in the country.[115]

Brazil's regulatory framework

- The "Brazilian Internet Bill of Rights" (MCI),[116] was 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."[117]
  - The bill underwent a long process of reviews, involving individuals, organisations, tech platforms, and other governments between 2009 and 2014.
  - The MCI makes Brazil one of the most populous countries in the world where the "democratic norm of equal access to information online" is inscribed in its Civil Code.
- Brazil's Anti-Terrorism law,[118] 2016, 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.

---

115. The first draft of the so-called "fake news" bill included provisions related to limiting the size of group chats and the possibility to forward messages, which were similar to current limitations used by WhatsApp.
116. Marco Civil da Internet
117. The principles of universality mainly relates to ensuring the diversity of internet users and "spurring innovation"
118. Lei nº 13.260, de 16 de Março de 2016

- The Brazilian Internet Freedom, Responsibility and Transparency Act,[119] or Law PLS2630/2020, known as the "fake news" bill:
  - The law was passed by the Senate in June 2020. However, no progress has been made since, and the law has yet to be approved by the Chamber of Deputies and signed by President Jair Bolsonaro – who is already expected to organise a popular consultation about the law, and ultimately veto it.[120]
  - The law is meant to counter the spread of misinformation online and would oblige messaging apps to implement measures to ensure the traceability of messages shared, as well as compel tech platforms to monitor "inauthentic behaviour".
  - The law does not set a territorial limit to its application, and instead would apply at "company level".

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:
- Judicial authorities have previously used defamation as a basis for ordering the removal of content.
- Authorities have also used violations of electoral laws as a ground for removal orders.
- Court orders can mandate the removal of content or blocking of accounts globally, as was the case with Facebook in August 2020 – see below.
- Brazil's "fake news" bill, if signed into a law, would have major consequences for online platforms, especially encrypted messaging services. This is because it will impose:
- Traceability requirements for messaging services, with messaging apps to be required to store the logs of "broadcasted messages" (meaning messages sent by more than 5 users and reaching at least a 1,000 users) for three months. This requirement is linked to a "technical capability directive" for platforms to be able to trace back individual messages.
- Verification of the identity of users upon receiving "reports of non-compliance with the fake news law, evidence of automated or inauthentic accounts, or upon court order". There is currently a lack of clarity regarding what is considered to be a "report of non-compliance" in this instance.
- The appointment of a local representative.
- Platforms employees'access to user databases in the country, in case they would be required to share user data with law enforcement.
- Additionally, the law could open the way for platforms liability for user-generated content.

---

119. Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet
120. See: Boadle Anthony (2020), Brazil's Bolsonaro would veto bill regulating fake news in current form, Reuters; and Tulio dos Santos Diogo (2021), Brazil, democracy, and the "fake news" bill, Global Americans.

TECH AGAINST TERRORISM COMMENTARY

Global application of the law

Tech Against Terrorism is concerned with the seemingly unlimited scope of the so-called "fake news" law, which is not limited to Brazil's jurisdiction, but would apply at "company level", no matter the user's location or nationality.[121] If the law is passed, Brazil's authority would have the power to regulate "fake news" online globally and request global communication to be traceable. A single country potentially setting the rules for online content worldwide represents a significant threat to freedom of expression online and users' rights.

This is not the first regulatory proposal with extra-territorial implications. The 2020 Rules in Pakistan are to apply to all online users of Pakistani nationality, no matter where they are located. Comparable to such mandates are judicial rulings, such as in Brazil, which compel tech companies to apply a removal or ban order worldwide rather than block access for local users. In August 2020, Facebook complied with a Brazilian's judge order to block the accounts of 12 of President Bolsonaro's supporters worldwide.[122] Facebook stated that it complied with the order due to the threat of criminal liability face by one of its employees.[123] In 2020, Facebook also had to comply with a court order to globally remove references to defamatory comments made against an Austrian politician.[124]

Content virality and traceability requirements

Similarly to India's 2021 Guidelines, Brazil is considering imposing a traceability requirement to counter the dissemination of fake news in the country. Unlike in India,[125] Brazil's traceability requirement is contingent on the "virality" of a message (1,000 users within 15 days).

---

121. Rodriguez Katitza and Schoen Seth (2020), <u>5 Serious Flaws in the New Brazilian "Fake News" Bill that Will Undermine Human Rights</u>, Electronic Frontier Foundation.
122. The individuals were under investigation for running a fake news network.
123. This is not the first time that a Facebook's employ faced criminal liability in Brazil due to the company not complying with a court order. In 2016, Diego Dzodan, Facebook's Vice President for Latin America was jailed for 24 hours, following a disputed court order for What's App to disclose user data for a drug-trafficking investigation.
124. And this until the injunction lasts, see: Lomas Natasha (2020), <u>Facebook loses final appeal in defamation takedown case, must remove same and similar hate posts globally</u>, TechCrunch.
125. In India, the traceability requirement is limited to "significant social media" and for certain investigatory or prosecution purposes.

However, this means that in practice platforms are to be able to trace back all messages given that any messages could become "viral". The Electronic Frontier Foundation criticised this as an infringement on due process, as it requires tech companies to retain logs of users' communication "before anyone has committed any legally defined offense".[126]

Furthermore, and as we raised in our commentary on the 2021 Guidelines in India, traceability requirements present major risks for online privacy and security. The technical changes that platforms must undergo in order to be able to retain information could weaken the end-to-end encryption offered by most messaging services. Tech Against Terrorism cautions any provisions that mandate platforms to modify their systems, and in particular their security protocols.

---

126. Rodriguez and Schoen (2020).

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.

The principle of net neutrality is enshrined in Colombia'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 be legally liable if they failed to adopt proper moderation mechanisms for comments published on their sites and online forums. According to the ruling, 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 enabled more restrictions on online spaces, and presented risks for online freedom of expression. On the Supreme Court's December 2019 decision, 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."[127]

---

127. "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."

In this same report, the FLIP identified a change towards the possibility of more stringent regulation of online platforms and content in Colombia. This is demonstrated by different legislative proposals, made between 2012 and 2019, that would have substantially limited freedom expression on the internet – and in some cases failed to meet the constitutional requirements in Colombia. Amongst the proposals underlined by FLIP is Bill 176/19 (2019) – which aimed to regulate the use of social media platforms.[128] The proposed bill outlined the need for requesting written consent to publish any type of information or data about a person, including photograph or videos. 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.

---

128. The proposal for Bill 176/19 did not succeed.
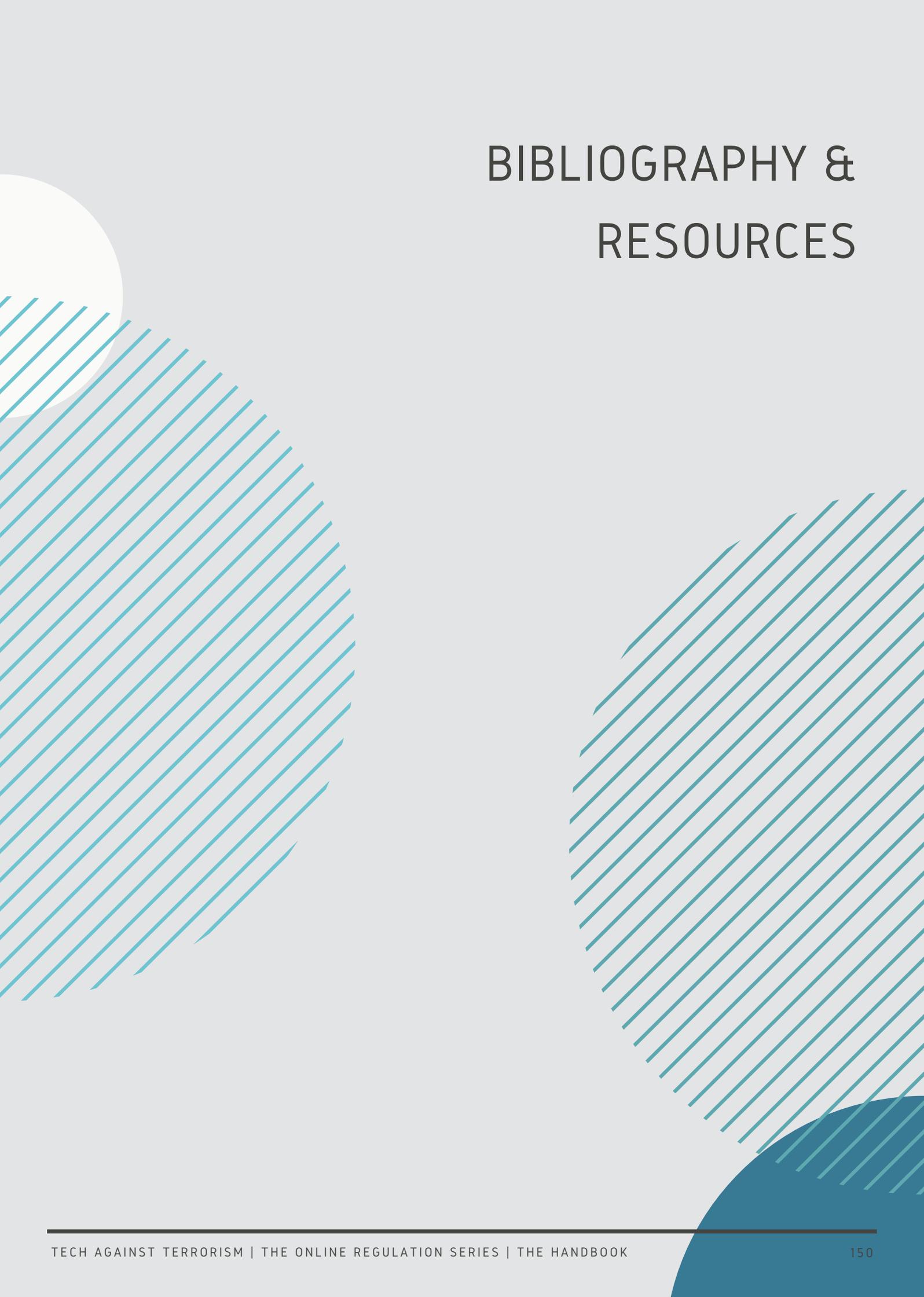
BIBLIOGRAPHY & RESOURCES

EXPERT RESOURCES

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

Aswad Evelyn Mary (2018), <u>The Future of Freedom of Expression Online.</u>

Balkin Jack (2019), <u>How to regulate (and not regulate) social media.</u>

Balkin Jack (2020), <u>How to Regulate (and not regulate) social media</u>, Yale Law School, Public Law Research Paper.

Barrett Paul M. (2020a), <u>Regulating Social Media: The Fight Over Section 230 — and Beyond</u>, NYU Stern.

Barrett Paul M. (2020b), <u>Why the Most Controversial US Internet Law is Worth Saving</u>, MIT Technology Review.

Benesch Susan (2020), <u>But Facebook's Not a Country: How to Interpret Human Rights et Human Rights Law for Social Media Companies.</u>

Caplan Robyn (2018), <u>Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches</u>, Data & Society.

Citron Danielle and Wittes Benjamin (2017), <u>The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity</u>, Fordham Law Review.

Citron Danielle (2020), <u>Section 230's Challenge to Civil Rights and Civil Liberties</u>, Boston Univ. School of Law, Public Law Research Paper.

Citron Danielle and Franks Mary Anne (2020), <u>The Internet As a Speech Machine and Other Myths Confounding Section 230 Reform</u>, Boston Univ. School of Law, Public Law Research Paper.

Douek Evelyn (2020), <u>Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability</u>, Columbia Law Review.

Douek Evelyn (2020), <u>The Limits of International Law in Content Moderation</u>, UCI Journal of International, Transnational, and Comparative Law (forthcoming 2021).

Kaye David (2018), <u>A Human Rights Approach to Platform Content Regulation.</u>

Kaye David (2019), <u>Speech Police: the Global Struggle to Govern the Internet.</u>

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

Keller Daphne (2018), <u>Internet Platforms: Observations on Speech, Danger, and Money</u>, Hoover Institution.

Keller Daphne, (2019), <u>Who Do You Sue?</u>, Hoover Institution.

Keller Daphne, (2020a), <u>Systemic Duties of Care and Intermediary Liability</u>, Stanford University Center for Internet and Society.

Keller Daphne (2020b), <u>Broad Consequences Of A Systemic Duty Of Care For Platforms</u>, The Center for Internet and Society, Stanford University.

Klonick Kate (2018), <u>The New Governors: The People, Rules, and Processes Governing Online Speech</u>, Harvard Legal Review.

Li Tiffany (2019), <u>Intermediaries and private speech regulation: a translatlantic dialogue – workshop report</u>, Boston University School of Law.

Lwin Michael (2020), <u>Applying International Human Rights Law for Use by Facebook.</u>

McDonald Stuart, Giro Correia Sara, and Watkin Amy-Louise (2019), <u>Regulating terrorist content on social media: automation and the rule of law</u>, International Journal of Law in Context.

McKelvey Fenwick, Tworek Heidi, Tenove Chris (2019), <u>How a standards council could help curb harmful content online</u>, Policy Options.

Sander Barrie (2020), <u>Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation.</u>

Zittrain Jonathan (2019), <u>Three Eras of Digital Governance.</u>

TECH SECTOR INITIATIVES

Article19 (2019), Social Media Councils: Consultation.

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

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

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

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

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

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

Dvoskin(2020), International Human Rights Law Is Not Enough to Fix Content Moderation's Legitimacy Crisis.

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

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

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

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

Perez Sarah (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 Courtney (2020), GIFCT: Possibly the Most Important Acronym You've Never Heard Of, JustSecurity.

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

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

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

Tech Against Terrorism (2020), The Terrorist Content Analytics Platform and Transparency By Design, VOX-Pol.

The Tech Against Terrorism Podcast (2020), Regulating the Online Sphere.

TikTok Newsroom (2020), Introducing the TikTok Asia Pacific Safety Advisory Council.

TikTok Newsroom (2021), Meet TikTok's European Safety Advisory Council.

Windwehr Svea and York Jillian (2020), One Database to Rule Them All, Vox-Pol.

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

COUNTRIES

SINGAPORE

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

Asia Internet Coalition (2020), <u>Toolkit – Addressing online misinformation through legislation</u>.

Asia One, <u>A look into Twitter's Asia-Pacific headquarters in Singapore.</u>

Chen Siyuan and Chia Chen Wei (2019),, Singapore Management University, Research Collection School of Law.

Chong Zoe (2018), <u>Facebook's Asia team moves to gigantic new headquarters in Singapore</u>, Cnet.

Consultancy.org (2020, <u>Singapore considered top alternative tech hub to Silicon Valley</u>.

EDB Singapore (2019), <u>Tech firms head to Singapore amidst Southeast Asia's growth.</u>

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

Infocomm – Media Development Authority (2016), <u>Internet Code of Practice</u>, Government of Singapore.

Infocomm – Media Development Authority, <u>Internet Regulatory Framework</u>, Government of Singapore.

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

Mullin Joe (2019), <u>Fancy New Terms, Same Old Backdoors: The Encryption Debate in 2019</u>, Electronic Frontier Foundation.

Newman Lily Hay (2020), <u>The EARN IT Act Is a Sneak Attack on Encryption</u>, Wired.

Pfefferkorn Riana (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 messages, The Guardian.

Ungku Fathin (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 Julia Carrie (2019), US, UK and Australia urge Facebook to create backdoor access to encrypted messages, The Guardian.

THE PHILIPPINES

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 Nick (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 Simon (2020), Digital 2020: The Philippines, DataReportal.

Khaliq Riyaz (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 Dona Z. (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 Marlon (2020), <u>AFP Dials down Facebook 'regulation'</u>, Inquirer.net.

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

Sen. Santiago Miriam D. (2012), <u>Magna Carta For Internet Freedom To Replace Anti-Cybercrime Law</u> Senate of the Philippines.

Yap Ben Dominic R., Protacio Jesus Paolo U., Lopez Jess Raymund M., and Lazaro Vladi Miguel S., (2020), <u>Anti-Terrorism Act signed into law</u>, Lexology.com.

York Jillian (2012), <u>Philippines' New Cybercrime Prevention Act Troubling for Free Expression</u>, Electronic Frontier Foundation.

AUSTRALIA

Arboleda Nico, (2020), <u>Telcos, Govt Reach Agreement on How to Block Terrorist Content</u>, CRN.

Baker McKenzie (2019),<u>Unprecedented Penalties for Enabling the Sharing of Abhorrent Violent Material Online</u>.

Baker McKenzie (2020), <u>Australian Government Opens Consultation on New Online Safety Act</u>.

Department of Communications and the Arts (2019), <u>Online Safety Legislative Reform Discussion Paper</u>, Australian Government.

Douek Evelyn (2019), <u>Australia's New Social Media Law Is a Mess</u>, Lawfare Blog.

Kaye David, Special Rapporteur on the Right to Freedom of Expression, Aolain Fionnula Ni, Special Rapporteur on Human Right and Fundamental Freedom while Countering Terrorism (2019), <u>Letter to the government of Australia</u>.

Hardy Keiran (2020), <u>Australia's encryption laws: practical need or political strategy?</u>, Internet Policy Review.

INDIA

Awashti Prashati (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.

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

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

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

Newton Casey (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 Saritha (2020), 400 million social media users are set to lose their anonymity in India, Bloomberg

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

Samuels Elyse (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; Awasthi Prashasti (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), <u>Use of the Internet for Terrorist Purposes</u>.

Wagner Kurt (2019), <u>WhatsApp is at risk in India. So are free speech and encryption</u>, Vox.

PAKISTAN

Amin Tahir (2021), <u>Procedure, Oversight and Safeguards 'Removal & Blocking of Unlawful Online Content Rules, 2021' modified</u>, Business Recorder.

Article19 (2016), <u>Pakistan: An Analysis of the Prevention of Electronic Crimes Bill 2015.</u>

Article19 (2020), <u>"Pakistan: Online Harms Rules violate freedom of expression"</u>

Asia Internet Coalition (2020), <u>"AIC Submits Response to Pakistan's Citizens Protection Rules (Against Online Harm)"</u>

Committee to Protect Journalists (2020), <u>"Pakistan government secretly passes strict social media regulations"</u>

CPU Media Trust (2020) <u>"Pakistan government secretly passes strict social media regulations"</u>

Digital Rights Foundation (2020a), <u>Citizens Protection (Against Online Harm) Rules, 2020: Legal Analysis.</u>

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

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

Farrell Nick (2020), <u>How Big Tech defeated Pakistan's censorship police.</u>

Garg Aryan (2020). <u>Pakistan's Online Harm Rules: Rights to Privacy and Speech Denied</u>, Jurist.org.

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

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

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

EUROPE

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

Article 19, Article 19's Recommendations for the EU Digital Services Act.

Citron Danielle (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 David, Ni Aoilain Fionnuala, Cannataci Joseph (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 Daphne (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 Adam & Berntsson Jacob (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.

FRANCE

Berne Xavier (2016), <u>Dans les coulisses de la plateforme de signalement Pharos</u>, NextInpact.

Breeden Aurelien (2020), <u>French court strikes down most of online hate speech law</u>, The New York Times.

Chandler Simon (2020),<u>France social media law is another coronavirus blow to freedom of speech</u>, Forbes.

Hadavas Chloe (2020), <u>France's New Online Hate Speech Law Is Fundamentally Flawed</u>, Slate.

La Maison des Journalistes, <u>Les limites de la liberté d'Expression.</u>

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

Lapowsky Issie (2020), <u>After sending content moderators home, YouTube doubled its video removals</u>, Protocol.

Lausson Julien (2020a), <u>Très contestée, la « loi Avia » contre la cyberhaine devient réalité</u>, Numerama.

Lausson Julien (2020b), <u>La loi Avia contre la haine sur Internet s'effondre quasi intégralement</u>, Numerama.

France Diplomatie (2019), <u>Réguler les contenus diffusés sur l'Internet et régulation des plateformes</u>, Ministère de l'Europe et des Affaires Etrangères.

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

Pielemeier Jason and Sheehy Chris (2019), <u>Understanding the Human Rights Risks Associated with Internet Referral Units</u>, The Global Network Initiative Blog.

Schulz Jacob (2020), <u>What's Going on With France's Online Hate Speech Law?</u>, Lawfare.

David Kaye (2019),<u>Mandat du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression.'</u>

GERMANY

Article 19 (2017), <u>Germany: Act to Improve Enforcement of the Law in Social Networks.</u>

Apple (2016), <u>A Message to Our Customers.</u>

Bayer Judith (2021), <u>Germany: New law against right-wing extremism and hate crime</u>, Inforrm.

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

Earp Madeline (2020). <u>Germany Revisits Influential Internet Law as Amendment Raises Privacy Implications</u>, Committee to Protect Journalists.

Echikson William (2020), <u>The Impact of the German NetzDG Law</u>, CEPS Europe.

Electronic Frontier Foundation (2016), <u>EFF to Support Apple in Encryption Battle</u>.

Kaye David (2019), <u>Speech Police: The Global Struggle to Govern the Internet: Columbia Global Reports.</u>

Grossman Lev (2016), <u>Inside Apple CEO Tim Cook's Fight With the FBI.</u>

Hardinghaus Alexander, Kimmich Romona & Schonhofen Sven (2020), <u>German Government Introduces New Bill to Amend Germany's Hate Speech Act, Establishing New Requirements for Social Networks and Video-Sharing Platforms,</u> Technology Law Dispatch, ReedSmith.

Heldt Amelie (2020), <u>Germany is amending its Online Speech Act NetzDg… but Not Only That</u>, Internet Policy Review.

Human Rights Watch (2018), <u>Germany: Flawed Social Media Law.</u>

Kahney Leander (2019), <u>The FBI Wanted a Back Door to the iPhone. Tim Cook Said No</u>, Wired.

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

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

Pielemeier Jason (2019), NetzDG: A Key Test for the Regulation of Tech Companies, GNI.

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

UNITED KINGDOM

Article 19, (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 Natasha (2019), UK Sets Out Safety-focused Plan to Regulate Internet Firms, Techcrunch

Osborne Clarke (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.

UK government (2020), Online Harms White Paper - Initial Consultation Response.

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

TURKEY

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

Beceni Yasin, Sevim Tuğrul, Aslan Erdem, Zengin Selen and Can Akdere Kaan (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), <u>Turkey: Internet Freedom, Rights in Sharp Decline.</u>

Human Rights Watch (2020), <u>Turkey: Press Freedom Under Attack.</u>

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

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

Zeldin Wendy (2014), <u>Turkey: Law on Internet Publications Amended</u>, Library of Congress.

UNITED STATES

Barrett Paul M. (2020a), <u>Regulating Social Media: The Fight Over Section 230 — and Beyond</u>, NYU Stern.

Barrett Paul M. (2020b), <u>Why the Most Controversial US Internet Law is Worth Saving</u>, MIT Technology Review.

Brody Jennifer, Null Eric (2020), <u>Unpacking the PACT Act</u>, AccessNow.

Feiner Lauren (2020), <u>GOP Sen. Hawley unveils his latest attack on tech's liability shield in new bill</u>, CNBC.

Future Tense (2021), <u>All the Ways Congress Wants to Change Section 230</u>, Slate.

Harmon Elliot, (2020), <u>It's Not Section 230 President Trump Hates, It's the First Amendment</u>, EFF.

Hawley Josh (2020), <u>Senator Hawley Announces Bill Empowering Americans to Sue Big Tech Companies Acting in Bad Faith.</u>

Mackey Aaron (2021), <u>Even with Changes, the Revised PACT Act Will Lead to More Online Censorship</u>, EFF.

Mullin Joe (2020), <u>Urgent: EARN IT Act Introduced in House of Representatives</u>, Electronic Frontiter Foundation.

Newton Casey (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 Alfred (2020), Why Your Privacy Could be Threatened by a Bill to Protect Children, CNET .

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

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

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

CANADA

Austen Ian (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 Alex, Oved Marco C., Silverman Craig, and L. Jane (2019, updated in 2020), Canadian government says it's considering regulating Facebook and other social media giants, The Hamilton Spectator.

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

Elghawaby Amira (2021), Canada is Bringing in New Legislation to Stop the Spread of Online Hate. Here's how it can work., Press Progress.

Leavitt Kierann (2021), After the Capitol riots, Ottawa draws lessons about social media regulation, Toronto Star.

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

Government of Canada (2021), Government of Canada takes action to protect Canadians against hate speech and hate crimes.

The Guardian (2019), <u>Canada may regulate social media companies to avoid election meddling</u>.

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

Innovation, Science and Economic Development Canada (2019), <u>Canada's Communications Future: time to act Broadcasting and Telecommunications Legislative Review</u>.

Jeftovic Marc E. (2020), <u>Canada's BTLR is a blueprint for regulating internet content</u>, Easydnscom.

Library of Congress, <u>Government Responses to Disinformation on Social Media Platforms: Canada</u>.

OpenMedia (2020), <u>The BTL...What? What is the BTLR report and what it means for the future of our Internet</u>.

Public Safety Canada (2019a), <u>Government of Canada Announces Initiatives to Address Violent Extremist and Terrorist Content Online</u>.

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

Silver Janet E. (2021) <u>Regulation of Online Hate Speech Coming Soon, Says Minister</u>, iPolitics.ca.

MOROCCO

Ait Akdim Youssef (2013), <u>Attentats de Casablanca : le 16 mai 2003, un « 11 septembre marocain</u>, Jeune Afrique.

Ait el Haj Rime (2013), <u>Le code numérique avorté!</u>, L'Economiste.

Article19 (2020), <u>Morocco: government must fully withdraw draft law on social media</u>.

APA News (2020), <u>Maroc : Un projet de loi sur l'utilisation des réseaux sociaux suscite l're de l'opinion publique</u>

Bouhrara Imane (2020), <u>Projet De Loi 22.20 : Le Gouvernement Tenterait-Il De Nous Faire Porter Des Muselières ?</u>, EcoActu.

LEconomiste (2020), <u>Loi 22.20: le projet de loi reporté jusqu'à la fin de l'urgence sanitaire</u>.

El Khamlichi Yasmine (2020), <u>Projet de loi 22.20 : Quid des droits du consommateur marocain ?</u>, Maroc Diplomatique.

Freedom House (2020), <u>Freedom on the Net: Morocco</u>.

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

Human Rights Watch (2020), <u>Morocco: Crackdown on Social Media Critics</u>.

Iraqi Fahd (2018), <u>Ce jour-là : 16 mai 2003, les attentats de Casablanca</u>, Jeune Afrique.

Maghraoui Abdeslam (2008), <u>Morocco's Reforms after the Casablanca Bombings</u>, Carnegie Endowment.

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

Oudrhiri Kaouthar (2020), <u>Projet de loi sur lutilisation des réseaux sociaux, un nouveau boulet pour lExécutif ?</u>, TELQUEL.

Perspective Monde, <u>16 Mai 2003: Attentats terroristes à Casablanca, au Maroc</u>, Sherbrook University, Quebec.

Zaireg Reda (2013), <u>Trois vérités toujours bonnes à rappeler sur le code numérique</u>, TELQUEL.

KENYA

Africa Times (2020) <u>Kenya warns of up to $50K fines for spreading fake COVID19 news.</u>

DLA Piper (2019) <u>Telecommunications Laws of the World: Kenya.</u>

Freedom House (2020), <u>Freedom on the Net 2020: Kenya</u>.

Freedom House (2018), <u>Online Survey: Kenyas Antiterrorism Strategy Should Prioritize Human Rights, Rule of Law</u>.

Hanibal Goitom (2014), <u>Kenya: Security Laws (Amendment) Bill Enacted,</u> Library of Congress.

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

Indokhomi Dominic, Syekei John (2020), THE COMPUTER MISUSE AND CYBERCRIMES ACT, Bowmans.

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

Munyua Alice, Githaiga Grace, Kapiyo Victor Intermediary Liability in Kenya, Association for Progressive Communications (APC).

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

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

UNODC, Kenya Training Manual on Human Rights and Criminal Justice Responses to Terrorism.

JORDAN

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. Department of Justice, United States Government.

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

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

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

Information and Research Center, A Glimpse into the Perception of Digital Privacy in Jordan. King Hussein Foundation.

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

Raed Omari (2018) <u>Jordanians launch social media campaign against new cybercrime law,</u> The Jordan Times.

Raed S.A. Faqir (2013) <u>Cyber Crimes in Jordan: A legal Assessment on the Effectiveness of Information System Crimes Law No (30) of 2010,</u> International Journal of Cyber Criminology .

Sevan Araz (2020). <u>Jordan Adopts Sweeping Cybersecurity Legislation.</u> The Middle East Institute.

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

BRAZIL

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

Al-Jazeera (2020), <u>Facebook bows to Brazil court order, bans pro-Bolsonaro profiles</u>.

Arnaudo Daniel (2017), <u>Brazil, the Internet and the Digital Bill of Rights: Reviewing the State of Brazilian Internet Governance</u>, Igarape Institute.

Boadle Anthony (2020), <u>Brazil's Bolsonaro would veto bill regulating fake news in current form, Reuters</u>.

Counter Extremism Project, <u>Brazil: Extremism & Counter-Extremism</u>.

Garcia Tsavkko Raphael (2020), <u>Brazil's "fake news" bill won't solve its misitionformation problem</u>

Isaac Mike and Roose Kevin (2018). <u>Disinformation Spreads on WhatsApp Ahead of Brazilian Election</u>, The New York Times.

Human Rights Watch (2015), <u>Brazil as the Global Guardian of Internet Freedom?</u>

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

Freedom House (2020), <u>Freedom on the Net: Brazil</u>.

Library of Congress (2016), <u>Brazil: New Anti-Terrorism Law Enacted</u>.

Lima Rafel (2020), <u>Misinformation campaigns escalate during Covid-19 pandemic in Brazil</u>.

Lomas Natasha (2020), <u>Facebook loses final appeal in defamation takedown case, must remove same and similar hate posts globally</u>, TechCrunch.

Lyons Kim (2020), <u>Brazil Supreme Court orders Facebook to block accounts of several Bolsonaro allies</u>, The Verge.

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

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

Rodriguez Katitza and Schoen Seth (2020), <u>5 Serious Flaws in the New Brazilian "Fake News" Bill that Will Undermine Human Rights</u>, Electronic Frontier Foundation.

Saraiva Augusta (2020), <u>Tackling Disinformation in Brazil</u>, Foreign Policy

Uchoa Pablo (2020), <u>Brazil coronavirus: 'Our biggest problem is fake news'</u>, BBC News.

COLOMBIA

Freedom House (2020), <u>Freedom on the Net: Colombia</u>.

Fundación Para la Libertad de Prensa (2019), <u>El internet que nadie quiere</u>,

Henshaw Alexis (2020), <u>Online extremism in Latin America – An Overview</u>, GNET.

Henshaw Alexis (2020), <u>Extremist responses to covid-19 in Latin America</u>, GNET.

Legis Ambito Juridico (2019), <u>Operadores de blogs pueden ser civilmente responsables por contenidos difamatorios</u>

Wired (2020), <u>What is Net Neutrality? The Complete WIRED guide</u>.

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

techagainstterrorism.org
@techvsterrorism
contact@techagainstterrorism.org

tech
against
terrorism