Module 3

Criminalisation of Online Speech

Advanced Modules on Digital Rights and Freedom of Expression Online

MEDIA DEFENCE

ISBN 978-0-9935214-1-6

Published by Media Legal Defence Initiative: www.mediadefence.org

This report was prepared with the assistance of ALT Advisory: https://altadvisory.africa/

November 2022

Table of Contents

MODULE 3

Criminalisation of Online Speech

The objectives of this module are:

- **To provide an overview of the criminalisation of online speech.**
- **To set out the applicable international human rights standards and fundamental international and regional legal principles.**
- **To understand the impact of criminalisation on freedom of expression and identify legitimate purposes for limiting freedom of expression.**
- **To set out and to examine the different forms of criminalisation, including hate speech, cybercrime and disinformation.**
- **To identify practical ways to deal with the competing interests of criminality and free speech.**

Introduction

There has been a growing trend of criminalising online speech over recent years. Many states have attempted to justify this as a response to threats of hate speech, national security, the mushrooming of cybercrimes, and the proliferation of disinformation. In many instances, this has led to the stifling of free speech and access to information. While some of the online harms that prompt criminalisation are a genuine concern which may warrant responses from states, there is an urgent need to ensure that states do not use these to justify restricting speech or controlling content.

This module provides an overview of the criminalisation of online speech. It looks at the applicable legal framework that guides what is permissible in terms of restrictions on the right to freedom of expression, and the relevant considerations for balancing competing rights. This module will also touch on hate speech, cybercrimes, and disinformation.

Overview of Criminalising Online Speech

Criminalisation, in the context of online speech, refers to the enactment of laws and policies that render specific forms of online expression illegal. Such criminalisation may be targeted at a range of harmful expressions, including:

- Hate speech;
- Threats or incitement to terrorism and violence;
- Disinformation;
- Defamation;

- Sexual abuse material including child sexual abuse material, the non-consensual dissemination of intimate images (**NCII**), and sexual exploitation online; and
- Cybercrimes.

From a criminal justice perspective, certain actions may warrant criminal consequences. However, in the context of online speech offences, there are a variety of competing considerations in the interplay between the offences, the rights they limit, and the limitations caused by creating the offences.

Walking the tightrope: criminalising online speech

The complexities of criminalising online speech should not be underestimated. The digital landscape, which in many ways has brought people together and facilitated free speech and dissent, has also created spaces that breed divisiveness, division, and exclusion. Supremacist ideologies, populist nationalism, gendered violence, and racism and xenophobia are some of the social ills that can take root in both our offline and online societies. Balancing dignity, equality, autonomy, and development against the right to free speech is not an easy task.

It is arguable that states' moves to impose restrictive measures on harmful speech, instead of addressing the systemic issues, such as the factors that enable the spread of misinformation online, are short-sighted solutions that restrict both those who are affected by online harms and those who are lawfully and legitimately expressing themselves. Organisations like the Collaboration on International ICT Policy in East and Southern Africa (**CIPESA**) and the Council of Foreign Relations (**CFR**) have noted with concern that governments the world over are adopting legislation that curtails free expression rights on the internet, either through the criminalisation of specific actions or through laws aimed at combating criminal activity online.

The right to freedom of expression is a fundamental human right that is protected in the Universal Declaration of Human Rights (**UDHR**), the International Covenant on Civil and Political Rights (**ICCPR**), and the African Charter on Human and People's Rights (**African Charter**). It is a right that is necessary for good governance and economic and social progress because it enables accountability by allowing people to freely debate and raise concerns with the government, including the protection and promotion of other human rights.[1]

Understanding the role of online speech offences, and their intended and unintended consequences requires careful navigation. Many laws that criminalise online speech are seen to be vague and overbroad and often fail to strike the appropriate balance between competing rights. These laws result in a chilling effect on the right to freedom of expression, whereby individuals steer clear of controversial topics because there is uncertainty about what is

[1] ARTICLE 19, 'Hate Speech' Explained: Toolkit (2015) (accessible at: https://www.article19.org/resources/hate-speech-explained-a-toolkit/).

permitted and what is not.[2] The chilling effect may be exacerbated where penalties for breach of the law are unduly harsh, as is the case with certain laws that criminalise online speech.

Applicable International Human Rights Standards

Overview of the right to freedom of expression and associated rights

It is trite that the right to freedom of expression is deeply entrenched as a fundamental human right and given protection through various international and regional instruments. Article 19 of the UDHR states:

> "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."

Article 19 of the ICCPR gives further effect to this, and article 20 of the ICCPR provides for certain restrictions on speech:

> "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."

In 2011, the United Nations Human Rights Committee published General Comment 34, which provides valuable guidance on how the right to freedom of expression should be interpreted. It states that freedom of expression is the "foundation stone for every free and democratic society", and that it is a "necessary condition for the realisation of the principles of transparency and accountability that are, in turn, essential for the promotion and protection of human rights." General Comment 34 notes that the right to freedom of expression includes:

- Political discourse.
- Commentary on one's own affairs and on public affairs.
- Canvassing ideas.
- Discussing human rights.
- Journalism.
- Cultural and artistic expression.
- Teaching, and religious discourse.

Freedom of expression may even extend to speech that may be regarded as deeply offensive by some people. The right applies both to verbal and non-verbal communications as well as all modes of expression, including audio-visual, electronic, and internet-based communication.

[2] Centre for Law and Democracy 'Restriction on freedom of expression' (accessible at: http://www.law-democracy.org/live/wp-content/uploads/2015/02/foe-briefingnotes-2.pdf)

Freedom of expression as an enabling right

- The United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression (**UNSR on FreeEx**) in a 2011 Report noted that the "right to freedom of opinion and expression is as much a fundamental right on its own accord as it is an 'enabler' of other rights". The UNSR went on to recognise that the right to freedom of expression also impacts economic, social, and cultural rights, such as the right to education and the right to take part in cultural life and to enjoy the benefits of scientific progress and its applications.

- General Comment 34 acknowledged that freedom of expression embraces the right of access to information, plays an important role in the conduct of public affairs, contributes to the effective exercise of the right to vote, and is integral to the enjoyment of the rights to freedom of assembly and association.

The 2017 Report of the UNSR on FreeEx sets out states' obligations under article 19 of the ICCPR. States may not interfere with, or in any way restrict, the holding of opinions, unless there are instances that warrant restriction – which must be provided by law and necessary for the respect of the rights or reputations of others, or for the protection of national security or public order, or public health or morals. States are also under an obligation to take steps to protect individuals from undue interference with human rights when committed by private actors, including taking appropriate steps to prevent, investigate, punish, and redress private actors' abuse. Such steps include the adoption and implementation of legislative, judicial, administrative, educative, and other appropriate measures that require or enable businesses to respect freedom of expression, and, where private sector abuses occur, access to an effective remedy.

In the African context, article 9 of the African Charter provides that:

> "1. Every individual shall have the right to receive information.
> 2. Every individual shall have the right to express and disseminate his opinions within the law."

African regional case law: limiting freedom of expression

- In *Constitutional Rights Project v Nigeria*, the African Commission on Human and People's Rights (**ACHPR**) held that the "only legitimate reasons for limitations of the rights and freedoms of the African Charter are found in Article 27(2), that is, that the rights "shall be exercised with due regard to the rights of others, collective security, morality and common interest". The ACHPR went on to state that the "justification of limitations must be strictly proportionate with and absolutely necessary for the advantages which follow. Most important, a limitation may not erode a right such that the right itself becomes illusory."

- The African Court on Human and People's Rights (**African Court**) in *Konaté v Burkina Faso* held that criminal sanctions for defamation must be necessary and proportionate, failing which they are incompatible with the ACHPR and other human rights instruments. Accordingly, expression must be within the prescripts of the law, and may only be limited in terms of article 27(2) of the African Charter, bearing in mind what is proportionate and necessary.

In 2002, the Declaration of Principles on Freedom of Expression in Africa was adopted to supplement article 9 of the African Charter. Article 2 of the Declaration of Principles established that arbitrary interference with a person's freedom of expression is prohibited and that any restrictions on freedom of expression shall be provided by law, serve a legitimate interest and be necessary in a democratic society. The revised 2019 Declaration of Principles on Freedom of Expression and Access to Information in Africa further provide that:

> "States shall criminalise prohibited speech as a last resort and only for the most severe cases. In determining the threshold of severity that may warrant criminal sanctions, States shall take into account the:
>
> a. prevailing social and political context;
> b. status of the speaker in relation to the audience;
> c. existence of a clear intent to incite;
> d. content and form of the speech;
> e. extent of the speech, including its public nature, size of audience
> f. and means of dissemination;
> g. real likelihood and imminence of harm."

It goes on to further state that States should not prohibit speech that merely lacks civility, or which offends or disturbs.[3]

Other implicated rights

In the context of online criminalisation, it is important to note that there are other interests and rights involved alongside the right to freedom of expression. These are different to the rights that are enabled through freedom of expression. The divergence of varying rights has been aptly captured in a 2019 Report on the UNSR on FreeEx:

> "[F]reedom of expression is a legal right of paramount value for democratic societies, interdependent with and supportive of other rights throughout the corpus of human rights law. At the same time, anti-discrimination, equality and equal and effective public participation underpin the entire corpus of human rights law. The kind of expression captured in article 20 of the International Covenant on Civil and Political Rights and article 4 of the International Convention on the Elimination of

[3] Principle 23 (3).

> All Forms of Racial Discrimination presents challenges to both sets of norms, something that all participants in public life must acknowledge."

Equality and non-discrimination are among the rights sometimes at odds with freedom of expression. While these rights can be exercised harmoniously, tensions are not uncommon. Beyond equality and non-discrimination, when considering freedom of expression and the criminalisation of online speech, regard should be had to other rights, including the rights of children. In some instances, protection measures online for children have at times taken a back seat to freedom of expression. In contrast, at other times, there have been constraints on children's or others' digital expression due to the need to combat online violence and exploitation. A 2017 UNICEF Report on Children's Rights and Business in a Digital World: Freedom of Expression, Association, Access to Information and Participation explains that what is ultimately required is some form of balancing between children's rights to freedom of expression and access to information and their right to be protected from violence.

There are other instances where there is also a need for balance:

- Balancing the right to freedom of expression with the right to privacy when determining whether to publish content.
- Striking a balance between the right to freedom of expression and the right to reputation.

It is necessary to note that rights are not absolute and may be subject to certain limitations and restrictions in order to balance competing rights and interests.[4] Ultimately, the right to freedom of expression is not unbounded and can be restricted to protect other rights, just as other rights may be subject to certain limitations and restrictions in order to advance freedom of expression. The restrictions of the right to freedom of expression will be dealt with further in the following section.

Restricting Freedom of Speech Online

As a result of the dramatic changes in the spread of information occasioned by the internet, there has been a proliferation of attempts to address issues relating to terrorism and national security, cybercrimes, and the spreading of disinformation online. Many of these attempts are, to varying degrees, in conflict with the right to freedom of expression.[5] Although the right to freedom of expression is a fundamental human right, it is not absolute. As with most rights, freedom of expression may be lawfully restricted where the restrictions are reasonable and justifiable in an open and democratic society. However, as confirmed in General Comment 34, the restrictions imposed by states should not put the right to freedom of expression in jeopardy.

Article 19(3) of ICCPR sets out the grounds upon which the right to seek, receive and impart information and ideas on the internet may be limited. Namely, the restriction must be:

[4] Media Defence, 'Training Manual on Digital Rights and Freedom of Expression Online Litigating digital rights and online freedom of expression in East, West and Southern Africa' at (accessible at https://www.mediadefence.org/resources/mldi-training-manual-digital-rights-and-freedom-expression-online).

[5] Shepard, 'Extremism, Free Speech and the Rule of Law: Evaluating the Compliance of Legislation Restricting Extremist Expressions with Article 19 ICCPR' Utrecht Journal of International and European Law (2017) (accessible at https://www.utrechtjournal.org/articles/10.5334/ujiel.405/).

- Provided by law.
- Necessary for respect for the rights of others, and for the protection of national security or of public order, or of public health or morals.

To determine whether a limitation of the right to freedom of expression is justifiable, a three-stage test must be applied in which it must be established that the limitation is:

- Provided by law.
- Pursues a legitimate aim.
- Necessary for a legitimate purpose.[6]

It is important to note that articles 19(3) and 20 of the ICCPR are compatible, and the prohibited grounds listed in article 20 can also be restricted in terms of article 19(3) and must also pass the three-stage test. It is further necessary to note that within the context of article 20, there is a need to recognise the distinction between protected and unprotected speech, and between what is prohibited and what is discriminatory, derogatory and demeaning discourse. Article 4(a) of the International Convention on the Elimination of All Forms of Racial Discrimination (**ICERD**) provides that certain forms of expression are prohibited and punishable by law. These include:

- Dissemination of ideas based on racial superiority or hatred.
- Incitement to racial discrimination.
- Acts or incitement of violence against any race or group of persons of another colour or ethnic origin of racially motivated violence.
- The provision of assistance, including of a financial nature, to racist activities.

Three-stage test for the justifiable criminalisation of online speech

The **first limb** (that the restriction is provided for by law) is relatively straightforward in relation to the criminalisation of online speech. The legislation must be clear, accessible, apply equally to everyone and be consistent with international human rights norms. Despite this, governments continue to enact laws that are vague, and which give themselves wide-ranging powers, including the power to decide what constitutes a legitimate purpose to restrict freedom of expression. On counter-terrorism measures, General Comment 34 provides that any offences relating to the encouragement of terrorism or extremist activity, or to the praising, glorifying, or justifying of terrorism, should be clearly defined to ensure that they do not lead to unnecessary or disproportionate interferences with freedom of expression. Excessive restrictions on access to information must also be avoided.

The **second limb** (that it pursues a legitimate aim) is more complicated and is important for the broader discussion on the criminalisation of online speech. In the current digital

[6] For a detailed outline of the limitation of freedom of expression see Module 2 on Restricting Access and Content at 4-5.

and political climate, the criminalisation of online speech is commonly used for political or other illegitimate purposes. Although there are legitimate grounds to restrict freedom of expression on the basis of national security, it is frequently subject to abuse.

The **third limb** requires an assessment of whether the restriction is necessary, where legislation provides for restricting freedom of expression for the legitimate purposes of protecting national security, countering terrorism, ensuring public order, or respecting the rights of others. In respect of necessity and proportionality, a 2019 Report of the UNSR on FreeEx notes that "restrictions must be demonstrated by the state as necessary to protect a legitimate interest and to be the least restrictive means to achieve the purported aim." A 2018 UNESCO report on world trends in freedom of expression and media development explains that this leg of the test can also cause controversy, when national security concerns are cited by states "to enact measures that present a clear challenge to media freedom, raising issues of necessity and proportionality." States are often quick to justify restrictions without fully considering the principle of necessity and whether less restrictive means are available. With new online threats, states are also becoming more restrictive, often in violation of the above test.

The different legitimate aims and the potential concerns that arise are discussed below.

National security

UNESCO has observed the growing trend of citing national security concerns as a justification for restricting freedom of expression. A legitimate national security interest is one that aims "to protect the existence of the nation or its territorial integrity or political independence against force or threat of force." This definition was laid out in the 1985 Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights. The Siracusa Principles further provide that a national security limitation:

> "cannot be invoked as a reason for imposing limitations to prevent merely local or relatively isolated threats to law and order" and "cannot be used as a pretext for imposing vague or arbitrary limitations and may only be invoked when there exists adequate safeguards and effective remedies against abuse."

The Johannesburg Principles on National Security, Freedom of Expression and Access to Information were drawn up in 1996 by a group of experts in international law, national security, and human rights. The principles state that in order for expression to be punished as a threat to national security, a government must show that:

- The expression is intended to incite imminent violence.
- It is likely to incite such violence.
- There is a direct and immediate connection between the expression and the likelihood or occurrence of such violence.

The Johannesburg Principles further provide that punishment (for disclosure of information) based on national security grounds is prohibited if the disclosure does not actually cause harm and is not likely to harm a legitimate national security interest.

The 2019 Declaration of the ACHPR further provides that "[f]reedom of expression shall not be restricted on public order or national security grounds unless there is a real risk of harm to a legitimate interest and there is a close causal link between the risk of harm and the expression."

Issues of national security have caused complications for the advancement of free expression for decades, including in the offline domain, as illustrated by the case note below.

Case note: Başkaya and Okçuoğlu v Turkey

In 1991, Mr Başkaya wrote a book which was published by Mr Okçuoğlu. Both Mr Başkaya and Mr Okçuoğlu are Turkish citizens. The book detailed the socio-economic revolution of Turkey and was critical of the ideology adopted by the state. The book came to the attention of the Turkish prosecution authorities, and Mr Başkaya was subsequently charged with disseminating propaganda against the indivisibility of the state. Mr Okçuoğlu was charged as the owner of the publishing company.

The National Security Court acquitted both men in 1992. However, the prosecutor subsequently successfully appealed the decision, which led to the matter being referred back to the trial court, which subsequently found both men guilty of the offences with which they had been charged. They were both sentenced to imprisonment and a fine. This decision was unsuccessfully appealed to the Court of Cassation, leading Mr Başkaya and Mr Okçuoğlu to approach the European Court of Human Rights (**ECtHR**).

Before the ECtHR, they argued, among other things, that their right to freedom of expression had been violated. The respondent state argued that the measures taken against the men were based on a law that was aimed at protecting interests such as territorial integrity, national unity, national security and the prevention of disorder and crime. The state further argued that they were convicted in pursuance of these legitimate aims since they had disseminated separatist propaganda vindicating the acts of the PKK (Workers' Party of Kurdistan), a terrorist organisation, which threatened these interests.

In 1999, the ECtHR delivered its decision, noting that freedom of expression is one of the "essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfilment", but may be subject to certain restrictions. The ECtHR emphasised that exceptions to freedom of expression must be construed strictly, and the need for any restrictions must be established convincingly. In conducting its limitations analysis, the ECtHR made the following observations:

- The requirement of "**necessary**" implies the existence of a "**pressing social need**".

- The content of the impugned statements and the context in which they were issued must be considered when determining if the interference was "**proportionate to the legitimate aims pursued**".
- Restrictions operate on a spectrum. There is little scope for restrictions on political speech or on debate on matters of public interest. However, there is a wider margin of appreciation when examining the need for an interference with freedom of expression in the context of remarks that incite violence.

The ECtHR, with due regard to Turkey's context, found that the measures taken by the state were in furtherance of the legitimate aim to ensure national security. However, the conviction and sentencing of Mr Başkaya and Mr Okçuoğlu was disproportionate to the aims pursued and therefore not "necessary in a democratic society". The ECtHR accordingly found that the right to freedom of expression had been violated.

National security can indeed be a legitimate aim; however, a restriction on freedom of expression must pass the other legs of the test as well, and cannot be justified on the legitimacy of national security grounds alone.

Case note: Good v Republic of Botswana

Mr Good, a political studies professor at the University of Botswana, was declared an undesirable inhabitant following the publication of a co-authored article which was critical of Botswana's presidential succession. Mr Good was deported without reason and was not provided with an opportunity to challenge the decision. After unsuccessfully exhausting all internal remedies, Mr Good approached the ACHPR, where he alleged that his right to be heard, his right to freedom of expression, his right to freedom of movement and his right to family life, all contained in the African Charter, had been violated.

In response to the allegation regarding the restriction of Mr Good's right to freedom of movement, the respondent state relied on national security as a justification, arguing that the ACHPR does not have competency over such issues as "[s]tates must be left alone and allowed to deal with matters of peace and national security". The respondent state did not address the alleged restriction on freedom of expression, and Mr Good argued that the respondent state failed to illustrate the nature of the so-called national security threat posed and why the deportation could be justified as proportionate in severity and intensity to the publication of the academic paper.

Despite the lack of a full response from the respondent state, the ACHPR analysed the alleged infringement and found that there is international consensus on the need to restrict freedom of expression for national security, but such a restriction must be **necessary**, serve a **legitimate interest** and be **provided for by law**. The ACHPR went on to note that notwithstanding the fact that "in the African Charter the grounds of limitation to freedom of expression are not expressly provided as in the other international and regional human rights treaties, the phrase 'within the law' under Article 9(2) provides a leeway to cautiously

fit in legitimate and justifiable individual, collective and national interests as grounds of limitation."

When conducting the limitations analysis, the ACHPR emphasised that a "higher degree of tolerance is expected when it is a political speech and an even higher threshold is required when it is directed towards the government and government officials." The ACHPR found that there was nothing in the article co-authored by Mr Good that could potentially create instability, unrest, or violence in the country; rather, it was merely the expression of opinions and views and did not amount to defamatory, disparaging, or inflammatory expression.

Ultimately, the ACHPR found that:

> "The action of the [r]espondent [s]tate was unnecessary, disproportionate and incompatible with the practices of democratic societies, international human rights norms and the African Charter in particular. The expulsion of a non-national legally resident in a country, for simply expressing their views, especially within the course of their profession, is a flagrant violation of [a]rticle 9(2) of the Charter."

Case note: SERAP v the Federal Republic of Nigeria

This case in the Community Court of Justice of the Economic Community of West African States (**ECOWAS**) also bears mention dealt with the Nigerian government's response to Twitter's removal of content tweeted by the President from its platform for violation of its rules.[7] Nigeria suspended the operations of Twitter arguing that its ongoing operations constituted threats to the stability of Nigeria and that "Twitter is undermining Nigeria's corporate existence" by allowing content that referred to separatist politics.

The ECOWAS Court emphasised the role of social media platforms such as Twitter as enablers of the rights to freedom of expression and access to information and held that the suspension was not made under any law or order of a court and that the government's mere reference or allusion to national security threats posed by protests in the country and their supposed potential to destabilise Nigeria did not constitute legal justification for the infringement on the right to freedom of expression.

As evinced in these cases, there are times when states will rely on national security when it is in fact not a legitimate aim. In such instances, courts should be quick to find the distinction between legitimate threats and critical expression.

[7] Media Defence and Mojirayo Ogunlana-Nkanga represented the applicants in this case.

Counter-terrorism

Terrorism and extremism, which are largely undefined and often misused terms, are frequently the basis for states to invoke restrictive measures on freedom of expression online in the name of national security. International human rights law provides extensive guidance for states on how to balance the real need to respond to terrorism, with the fundamental right to freedom of expression.

The 2015 Joint Declaration on Freedom of Expression and Responses to Conflict Situations by Special Rapporteurs on Freedom of Expression provides that:

> "States should refrain from applying restrictions relating to 'terrorism' in an unduly broad manner. Criminal responsibility for expression relating to terrorism should be limited to those who incite others to terrorism; vague concepts such as glorifying', 'justifying' or 'encouraging' terrorism should not be used."

The 2016 Joint Declaration on Freedom of Expression and Countering Violent Extremism notes that:

> "Everyone has the right to seek, receive and impart information and ideas of all kinds, especially on matters of public concern, including issues relating to violence and terrorism, as well as to comment on and criticise the manner in which [s]tates and politicians respond to these phenomena."

The 2016 Joint Declaration further provides that states are obliged to ensure that there is an enabling environment for the media to keep society informed, "particularly in times of heightened social or political tensions". This point is also emphasised in General Comment No. 34 on the ICCPR, which states that the media plays an important role in informing the public about acts of terrorism and that it should be able to perform its legitimate functions and duties in this regard without hindrance.[8]

At a minimum, if there is to be a limitation of access to the internet or to content online on the grounds of anti-terrorism, there should be transparency regarding the laws, policies and practices relied upon, clear definitions of terms such as 'national security' and 'terrorism', and independent and impartial oversight being exercised.

There is also a general presumption in international law that prior restraint – restricting access to content before it has been published – is unnecessary and disproportionate. Although there may be a strong argument for the need to step in to stop the dissemination of information prior to publication of content relating to terrorism, the courts have stressed that prior restraint can only be allowed in exceptional circumstances and must be robustly justified.[9]

[8] UN Human Rights Council, 'General Comment no. 34 at para 46 (2011) (accessible at https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf).
[9] For example, see *New York Times Co. v United States (1971)* and *Amnesty International Togo and Ors v. The Togolese Republic.*

Public order offences

Public order offences can be developed and implemented to provide for legitimate aims, especially in the context of security forces. This means that laws which allow security forces to limit free speech to protect public order may be legitimate, as long as they comply with the requirements listed above. This legitimate aim is one that should not be abused due to the significant impact it can have on the people affected by the restriction on freedom of speech. This is particularly evident in the recent proliferation of internet shutdowns during crucial election periods. These acts are usually commissioned under the guise of maintaining public order, whereas they constitute an effort by states to silence dissent. The consequences of internet shutdowns are that the public's right to access information, which may be crucial at a particular time, is violated.[10] For more on internet shutdowns, see Advanced Module 2 of this series on Restricting Access and Content.

UNESCO Training Modules on Public Order and Freedom of Expression

In response to tensions between the maintenance of public order and the restrictions on freedom of expression, particularly in the context of journalism, UNESCO has developed training modules to empower both security forces and journalists to understand the law and their respective roles and responsibilities.

- The 2015 Freedom of Expression and Public Order Training Manual provides legal references and tools for security forces to promote transparency, facilitate and improve relations between security forces and the media, and encourage respect for the safety of journalists in the field.

- The 2018 report from UNESCO Freedom of Expression and Public Order: Fostering the Relationship between Security Forces and Journalists seeks to facilitate the relationship between security forces and journalists in order to establish an enabling environment for journalists. This training manual aims to empower journalists and citizens in order for them to exercise their rights to freedom of expression and access to information. It focuses on the importance of transparent law enforcement institutions, which respect freedom of expression and the right to information and promote accountability and the rule of law while respecting human rights.

- In 2022, UNESCO, together with the International Police Association and IBZ Gimborn Castle launched a joint Massive Open Online Course (**MOOC**) for members of law enforcement and police officers to raise awareness of international and regional standards on freedom of expression, access to information, and safety of journalists.

[10] For more on internet shutdowns see Module 2 on Restricting Access and Content.

Forms of Criminalisation

In 2019, the ACHPR recognised that the primary issues relating to freedom of expression include:

- Co-regulation of the media.
- Safety of journalists.
- Restrictions related to cyber-crime laws.
- Regulation of the internet.

While there is an array of actions and forms of speech that have attracted criminal sanctions, this section focuses on hate speech, cybercrimes, and disinformation.[11]

Hate speech

The reconciliation of values

A 2019 Report by the UNSR on FreeEx found that:

> "Under international human rights law, the limitation of hate speech seems to demand a reconciliation of two sets of values: democratic society's requirements to allow open debate and individual autonomy and development with the compelling obligation to prevent attacks on vulnerable communities and ensure the equal and non-discriminatory participation of all individuals in public life. Governments often exploit the resulting uncertainty to threaten legitimate expression, such as political dissent and criticism or religious disagreement."

The above statement of the UNSR illustrates some of the complexities regarding the criminalisation of hate speech. The escalation of prejudice and intolerance has led many governments to criminalise hate speech. However, this creates inherent difficulties because hate speech is a vague term that lacks universal understanding, and such provisions are open to abuse and restrictions on a wide range of lawful expression.

Overview of international instruments dealing with hate speech

- Article 20(2) of the ICCPR obliges states to prohibit by law "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."

- The Rabat Plan of Action was introduced in 2012 to provide recommendations on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. It outlines six factors that should be considered when determining whether a speaker intends and is capable of having the effect of

[11] For more on specific types of speech-related offences, see Media Defence above n 3 at 48-61.

inciting their audience to engage in violent or discriminatory action through the advocacy of discriminatory hatred.

The 2019 report of the UNSR on FreeEx also evaluates the human rights law that applies to the regulation of online hate speech and recommends that States should not treat online hate speech as a separate category from offline hate speech with higher penalties, should strictly define what constitutes prohibited content in their domestic laws, and should resist criminalising speech except in the gravest situations.

Rabat Plan of Action: Six-part threshold test for expressions considered criminal offences

"**Context**: Context is of great importance when assessing whether particular statements are likely to incite discrimination, hostility or violence against the target group, and it may have a direct bearing on both intent and/or causation. Analysis of the context should place the speech act within the social and political context prevalent at the time the speech was made and disseminated.

Speaker: The speaker's position or status in society should be considered, specifically the individual's or organization's standing in the context of the audience to whom the speech is directed.

Intent: Article 20 of the International Covenant on Civil and Political Rights anticipates intent. Negligence and recklessness are not sufficient for an act to be an offence under article 20 of the Covenant, as this article provides for "advocacy" and "incitement" rather than the mere distribution or circulation of material. In this regard, it requires the activation of a triangular relationship between the object and subject of the speech act as well as the audience.

Content and form: The content of the speech constitutes one of the key foci of the court's deliberations and is a critical element of incitement. Content analysis may include the degree to which the speech was provocative and direct, as well as the form, style, and nature of arguments deployed in the speech, or the balance struck between arguments deployed.

Extent of the speech act: Extent includes such elements as the reach of the speech act, its public nature, its magnitude, and the size of its audience. Other elements to consider include whether the speech is public, what means of dissemination are used, for example, a single leaflet or broadcast in the mainstream media or the Internet, the frequency, quantity, and extent of the communications, whether the audience had the means to act on the incitement, whether the statement (or work) is circulated in a restricted environment or widely accessible to the general public.

Likelihood, including imminence: Incitement, by definition, is an inchoate crime. The action advocated through incitement speech does not have to be committed for said speech to amount to a crime. Nevertheless, some degree of risk of harm must be identified. It means that the courts will have to determine that there was a reasonable probability that the speech would succeed in inciting actual action against the target group, recognising that such causation should be rather direct."

Identifying hate speech

It is sometimes tricky to distinguish between speech that is protected and that which constitutes hate speech.

- Hate speech may be prohibited only if the prohibition meets the standards of article 19(3), namely:[12]
    - **Legality**: laws criminalising hate speech must be precise, public, and transparent.
    - **Legitimacy**: laws should be justified to protect and respect the rights or reputations of others or to protect national security, public order, public health or morals.
    - **Necessity and proportionality**: the criminalising legislation must protect a legitimate interest and be the least restrictive means to achieve the purported aim.
- Hate speech is lawful and should be protected if it is:
    - Inflammatory or offensive expression that does not meet the above thresholds. Notably, this may include speech that is critical or that causes shock or offence.

Online hate speech

The nature of online domains, such as social media, creates conditions for the sharing and spreading of hate speech that are relevant to considerations of how to appropriately regulate hate speech. For example:

- Content is more easily posted online without due consideration or thought. Regulation must distinguish between poorly considered statements posted hastily online, and an actual threat that is part of a systemic campaign of hatred.
- Once something is online, it can be difficult (or impossible) to get it off entirely. Hate speech posted online can persist in different formats across multiple different platforms, which can make it difficult to police.
- Online content is frequently posted under the cover of anonymity, which presents an additional challenge to dealing with hate speech online.

[12] Article 19, 'Hate Speech Explained: A Toolkit,' (2015) (accessible at: https://www.article19.org/data/files/medialibrary/38231/'Hate-Speech'-Explained---A-Toolkit-%282015-Edition%29.pdf).

- The internet has transnational reach, which raises cross-jurisdictional complications in terms of legal mechanisms for combatting hate speech.

ARTICLE 19 Hate Speech Explained: A Toolkit

ARTICLE 19 has published a toolkit on identifying and countering hate speech while protecting the rights to freedom of expression and equality. The toolkit responds to a growing demand for clear guidance on identifying 'hate speech' and for responding to the challenges hate speech poses within a human rights framework.

It is clear that cooperation from the state can be an effective means of safeguarding human rights. However, states are not always fulfilling their duties. Accordingly, lawyers, civil society organisations (**CSOs**), individuals, and community members need to work together to ensure that states are acting in compliance with their international human rights obligations. This can include strategic litigation, policy reform and advocacy, such as:

- Ensuring that states are creating an **enabling environment** for the right to freedom of expression. This can include ratifying international and regional human rights instruments, adopting domestic laws to protect freedom of expression and repealing any laws that unduly limit the right to freedom of expression.
- Ensuring that states **safeguard** the rights of individuals who exercise their right to freedom of expression. This requires ensuring that states make a concerted effort to end impunity for attacks on independent and critical voices.
- Ensuring that **domestic laws** guarantee equality before the law and equal protection of the law. That includes protection against discrimination on all grounds recognised under international human rights law.
- Ensuring that states establish or strengthen the role of **independent equality institutions** or expand the mandate of national human rights institutions.
- Ensuring that states adopt a **regulatory framework** for diverse and pluralistic media, which promotes pluralism and equality.

Some of these elements of online hate speech were addressed in recent cases in South Africa:

- In *South African Human Rights Commission (**SAHRC**) v Matumba,* involving the posting of hate speech online by a person allegedly using a false account on a social media platform, the Equality Court in South Africa considered whether a series of tweets posted in 2020 constituted harassment in terms of the country's law protecting equality. The SAHRC argued that the tweets included "serious, demeaning and humiliating comments against women, and black women in particular". An *amicus curiae* brief submitted by Media Monitoring Africa (**MMA**), a civil society organisation, highlighted the speed and application of content on Twitter, as opposed to more traditional formats, and analysed how to determine the reasonable reader in the context of social media.

- Another case in South Africa provided a detailed analysis of the line between "hurtful" and "hate" speech. In *Qwelane v. South African Human Rights Commission*, the South African Constitutional Court held that the prohibition of "hurtful" speech was an unjustifiable infringement of the right to freedom of expression, but held that the hate speech provision could be made constitutional by limiting it to expression that was intended *to be harmful or incite harm and to promote or propagate hatred.*

Incidences of hate speech regulation

Unfortunately, there are numerous examples of countries attempting to pass, or successfully passing, hate speech legislation that includes criminal penalties, particularly in Africa:

- In 2020 Ethiopia enacted the Hate Speech and Disinformation Prevention and Suppression Proclamation which, while having seemingly well-intentioned objectives, has been decried by civil society as a threat to freedom of expression and access to information online.[13]

- In Nigeria, the National Commission for the Prohibition of Hate Speech Bill was tabled in 2019 which would prohibit "abusive, threatening, and insulting behaviour", and another bill under consideration in 2022 proposes to classify hate speech as an electoral offence that may attract a jail term of 10 years or a fine of N40m or both.

- In South Africa, a bill on the prevention of hate crimes and hate speech has been criticised for its potential to be used to silence free speech and criticism and to stymie difficult discussions about race, gender, and sexuality.

Cybercrime

There is no single uniform or universally accepted definition for cybercrime, and there is an ongoing debate as to what the term entails. Some of the explanations and definitions advanced cover "a whole slew of criminal activity" including the theft of personal information, fraud, and the dissemination of ransomware.[14] Cybercrimes can also be the online extension of existing offline crimes such as harassment and sexual abuse, or producing, offering to make available, or making available, and distributing racist and xenophobic material.[15] For ease of reference, cybercrimes may be categorised as follows:[16]

[13] CIPESA, Edrine Wanyama, 'Ethiopia's New Hate Speech and Disinformation Law Weighs Heavily on Social Media Users and Internet Intermediaries' (2020) (accessible at: https://cipesa.org/2020/07/ethiopias-new-hate-speech-and-disinformation-law-weighs-heavily-on-social-media-users-and-internet-intermediaries/).
[14] Microsoft, 'Cybercrime and freedom of speech – a counterproductive entanglement' (2017) (accessible at https://www.microsoft.com/security/blog/2017/06/14/cybercrime-and-freedom-of-speech-a-counterproductive-entanglement/).
[15] See UNODC, 'Module 2: General Types of Cyber Crime; E4J University Module Series: Cybercrime (2019) (accessible at https://www.unodc.org/e4j/en/cybercrime/module-2/key-issues/intro.html) and UNODC 'Module 3: Legal Frameworks and Human Rights' E4J University Module Series: Cybercrime (2019) (accessible at https://www.unodc.org/e4j/en/cybercrime/module-3/key-issues/international-human-rights-and-cybercrime-law.html).
[16] Id. See further ITU 'Understanding cybercrime: Phenomena, challenges and legal response' (2012) (accessible at http://www.itu.int/ITU-D/cyb/cybersecurity/docs/Cybercrime%20legislation%20EV6.pdf).

Category	Examples of crimes
Offences against the confidentiality, integrity and availability of computer data and systems	Illegal access (hacking) • Password breaking • Distributed denial-of-service (DDoS) attacks • Automated attacks and botnets
	Illegal data acquisition (data espionage) • Scanning for unprotected ports • Circumventing protection measures • Social engineering • Phishing
	Illegal interception • Intercepting communications to record the information exchanges • Setting up fraudulent access points
	Data interference • Deleting, suppressing, or altering computer data • Creation of malware and computer viruses
Content-related offences	• Sexual exploitation material • Child sexual abuse material • Commercial sexual exploitation of children • Racist and xenophobic speech, hate speech and promotion of violence • Disinformation and fake news
Copyright and trademark-related offences	• Reproduction of material • Exchange of copyright-protected material (songs and movies) • Certain file-sharing systems • Domain name-related offences
Computer-related offences	• Computer-related fraud • Online auction fraud • Advance fee fraud • Identity theft • Cyberstalking, cyberharassment, and cyberbullying

Cybercrime and cybersecurity are two interlinked issues in an interconnected digital environment. Cybersecurity, or the management and prevention of cybercrime, refers to the collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance, and technologies that can be used to protect the cyber-environment and organisational and user's assets, such as computing devices, applications, and telecommunication systems.[17]

[17] ITU Definition of Cybersecurity, (accessible at: https://www.itu.int/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx).

Overview of international instruments

Currently, there are three prominent international instruments that engage the topic of cybercrime:[18]

- The 2001 Convention on Cybercrimes (Budapest Convention) is the first international treaty that seeks to address internet and computer crimes. Its main objective is to pursue a common criminal policy aimed at the protection of society against cybercrimes by adopting appropriate legislation and fostering international co-operation.

- The Additional Protocol to the Convention on Cybercrimes concerns the criminalisation of acts of a racist and xenophobic nature committed through computer systems. As an international legal instrument, the Protocol provides guidance and plays a key role in facilitating harmonisation across different legal regimes on the issue of specific forms of online speech.

- The 2014 African Union Convention on Cybersecurity and Personal Data (Malabo Convention), is a treaty dealing with cybercrime, data protection and related issues for the African continent. As of 2022, the Malabo Convention was still undergoing ratification, but if brought into force the Convention, among other things, encourages states to take necessary legislative and/or regulatory measures to establish criminal offences relating to cybercrimes. The offences include:

    - Creating, downloading, disseminating, or making available in any form writings, messages, photography, drawings or any other presentation of ideas or theories of racist or xenophobic nature through a computer system.
    - Threatening, through a computer system, to commit a criminal offence against a person for the reason that they belong to a group distinguished by race, colour, descent, national or ethnic origin or religion, where such membership serves as a pretext for any of these characteristics.
    - Insulting, through a computer system, persons for the reason that they belong to a group distinguished by race, colour, descent, national or ethnic origin or religion or political opinion if used as a pretext for any of these factors, or against a group of persons distinguished by any of these characteristics.

Under the Malabo Convention, states are also urged to enact legislation criminalising acts related to child pornography. Importantly, the Malabo Convention does identify acts that warrant criminalisation, such as child pornography and racist and xenophobic acts. However, there are some concerns when it comes to free speech in the online context. For instance, the Malabo Convention uses vague language which may be open to abuse by states. An example is the provision that criminalises the use of **insulting language,** which is problematic because it describes a significant portion of the language used on the internet. This can lead to subjective prosecutions and, eventually, may lead to criminal convictions for what should be

[18] Global Action on Cybercrime Extended, 'Comparative analysis of the Malabo Convention of the African Union and the Budapest Convention on Cybercrime' (2016) (accessible at https://rm.coe.int/16806bf0f8).

protected speech. The Convention also raises concerns in that it expands the search and seizure powers of the state.

The rise in cybercrime laws

The UNODC has found that cybercrime laws are of particular relevance to the criminalisation of online speech because the laws that are enacted to regulate cybercrimes can result in the restriction of freedom of expression. Access Now notes that one of the main concerns about the plethora of laws that are currently being enacted to regulate cybercrimes is that many of them lack clear definitions and are susceptible to being used to regulate online content and restrict freedom of expression. This is a growing concern among human rights defenders as many have been subjected to a wave of arrests and convictions in what is an escalating assault on freedom of expression through cybercrime laws.

Cybercrime laws in Nigeria

While there may be legitimate aims in enacting these laws, there are serious concerns that many of these laws are vague and overbroad and are susceptible to being used to restrict freedom of expression. Amnesty International has reported a growing trend of arrests, detention and torture of journalists and bloggers as well as pointed attacks on major media houses. Journalists and bloggers are reportedly being charged with cybercrimes under Nigeria's Cybercrime Act, which criminalises a substantial number of online forms of expression.

This situation may be exacerbated if the proposed Protection from Internet Falsehoods and Manipulation Bill is passed into law. The Bill is aimed at enabling measures to be taken to detect, control and safeguard against uncoordinated and inauthentic behaviour and other misuses of online accounts and bots, enabling measures to be taken to enhance disclosure of information regarding paid content directed towards a political end and to sanction offenders.

The Bill seeks to criminalise, among other things, prohibited statements of facts which include false statements of fact and statements that are likely to be prejudicial to the country's security, public health, public safety, public tranquillity or finances, prejudice Nigeria's relations with other countries, influence the outcome of an election or referendum, incite feelings of enmity, hatred towards a person, ill will between a group of persons, or diminish public confidence in the performance or exercise of any duty, function or power by the government.

If this Bill is passed it could mean a further affront to freedom of expression in Nigeria, which as it stands is under threat due to the cybercrime legislation that is already in existence. Further, the Bill gives the State wide-ranging powers, which may be susceptible to abuse.[19]

[19] For further commentary on trends in Africa see CIPESA, 'Why are African Governments Criminalising Online Speech? Because They Fear Its Power' (2018) (accessible at https://cipesa.org/2018/10/why-are-african-governments-criminalising-online-speech-because-they-fear-its-power/).

The Government of Nigeria's extended suspension of the operations of the social networking service, Twitter, for close to six months up to January 2022 appear to have reignited the government's interest to implement the proposed Bill.
It is further worth noting that in 2020, the ECOWAS Community Court of Justice ordered Nigeria to repeal its cybercrime legislation, which was held to violate the right to freedom of expression.

In relation to the concerns regarding cybercrime legislation, a 2019 Report of the UNSR on FreeEx noted:

> "A surge in legislation and policies aimed at combating cybercrime has also opened the door to punishing and surveilling activists and protesters in many countries around the world. While the role that technology can play in promoting terrorism, inciting violence and manipulating elections is a genuine and serious global concern, such threats are often used as a pretext to push back against the new digital civil society."

UN Resolution on Countering the Use of Information and Communications Technologies for Criminal Purposes

In July 2019, the United Nations General Assembly presented a Draft Resolution on countering the use of information and communications technologies for criminal purposes.

CSOs were highly critical of the resolution, calling for delegations to vote against it. In an Open letter to UN General Assembly, the following concerns were raised:

- The "use of information and communications technologies for criminal purposes" is not defined in the resolution, which is not just a concern from an accuracy perspective; but also opens the door to criminalising ordinary online behaviour that is protected.
- While legislation aimed at addressing cybercrime can be necessary and reinforce democratic institutions, when misused, cybercrime laws can create a chilling effect.
- It goes far beyond what the Budapest Convention allows for regarding cross-border access to data, including by limiting the ability of a signatory state to refuse to provide access to requested data.
- Building on and improving existing instruments is more desirable and practical than diverting already scarce resources into the pursuit of a new international framework, which is likely to stretch over many years and unlikely to result in consensus.
- The establishment of an ad hoc intergovernmental committee of experts to address the issue of cybercrime would exclude key stakeholders who bring valuable expertise and perspectives.

Despite these concerns, the resolution was adopted and published in January 2020. Through the resolution, an open-ended ad hoc intergovernmental committee of experts, representative

of all regions, will be established to elaborate a comprehensive international convention on countering the use of ICTs for criminal purposes, taking into full consideration existing international instruments and efforts at the national, regional, and international levels on combating the use of ICTs for criminal purposes.

Lawyers and activists should monitor further developments in relation to this and, where possible, engage with relevant stakeholders in order to positively influence future developments and decisions.

Disinformation and 'fake news'

Disinformation includes statements which are known or reasonably should be known to be false that seek to mislead the public, and, in turn, interfere and inhibit the ability of the public to seek, receive, and impart information.[20] In 2018, the High-Level Expert Group on Fake News and Online Disinformation defined disinformation to mean—

> "all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit. It does not cover issues arising from the creation and dissemination online of illegal content. Nor does it cover other forms of deliberate but not misleading distortions of facts, such a satire and parody."

Disinformation that is designed to look like news content is sometimes popularly referred to as "fake news." The High-Level Expert Group noted two reasons for avoiding the use of this term:

- The term is inadequate to capture the complex problem of disinformation, which involves content that blends fabricated information with facts.
- The term is misleading as it has been appropriated by some politicians and their supporters to dismiss coverage that they find disagreeable and has thus become a weapon with which powerful actors can interfere in the circulation of information and attack and undermine independent news media.

Concerted disinformation campaigns by foreign state and non-state actors to interfere in the 2016 US presidential elections brought unprecedented light on the issue of "fake news" and the ease with which disinformation can be disseminated online.[21] The COVID-19 pandemic also highlighted the capacity for the rapid spread of disinformation, which undermined efforts to address the disease and roll-out treatments and vaccines.

In response to this growing trend of disinformation, a number of states have enacted legislation criminalising the online publication of false statements. Such responses continue to increase in speed and magnitude and to cause demonstrable and significant public harm. The 2017

[20] Access Now, Civil Liberties Union for Europe and European Digital Rights 'Informing the disinformation debate' (2018) (accessible at https://dq4n3btxmr8c9.cloudfront.net/files/2r7-0S/online_disinformation.pdf).
[21] Vox, '4 main takeaways from new reports on Russia's 2016 election interference' (2019) (accessible at: https://www.vox.com/world/2018/12/17/18144523/russia-senate-report-african-american-ira-clinton-instagram)

Joint Declaration on Fake News, Disinformation and Propaganda by the UN Special Rapporteur on Freedom of opinion and expression together with his counterparts from the Organization for Security and Co-operation in Europe (**OSCE**), the Organization of American States (**OAS**), and the African Commission on Human and Peoples' Rights (**ACHPR**), noted that countering these issues poses complex challenges that could result in censorship and the suppression of critical thinking.

Addressing fake news

Various international bodies, states and organisations have grappled with different responses to the complexities of disinformation. However, many countries have responded with harsh legislation that does not strike an appropriate balance between addressing disinformation and protecting the right to freedom of expression. The advent of the COVID-19 pandemic and associated disinformation has further accelerated this trend. Some examples include:

- **Malaysia**: In 2018, the Malaysian government enacted the Anti-Fake News Act, which attaches criminal liability to persons who knowingly create, offer, publish, print, distribute, circulate, or disseminate fake news. The Act defined "fake news" as including "any news, information, data and reports, which is or are wholly or partly false, whether in the form of features, visuals or audio recordings or in any other form capable of suggesting words or ideas."[22] However, the existence of the Act was short-lived. It was repealed by the Anti-Fake News (Repeal) Act 825 of 2020, with government citing its commitment to abolish draconian laws and protect media freedom. However, in March 2021, the government issued the Emergency (Essential Powers) (No. 2) Ordinance 2021 which criminalises the dissemination of fake news related to COVID-19 and repeated many of the problematic provisions of the Anti-Fake News Act.

- **Cameroon:** The Penal Code in Cameroon criminalises the sending out or propagation of false information. Section 113 imposes a penalty of imprisonment between three months to three years and a fine between CFAF 100 000 (approximately USD172) to CFAF 2 000 000 (approximately USD3400) for persons found guilty of this offence. The Committee to Protect Journalists (**CPJ**) has noted with concern the arrest and detention of journalists under this provision, in particular, a journalist who was sent to maximum-security prison on charges of defamation and spreading false news.

- **Russia**: In 2019, the State Duma (the Russian Federal Assembly) passed legislation on Information, Information Technologies and Protection of Information, and a Code of Administrative Offenses both aimed at countering "fake news". ARTICLE 19 explains that these amended laws allow authorities in Russia to block websites that they consider to be publishing disinformation. Websites are also liable for insulting Russian authorities. The Moscow Times reported that "online news outlets and users that spread "fake news" will face fines of up to 1.5 million Rubles (USD20 000) for repeat offences. Insulting state symbols and the authorities, including Vladimir Putin, will carry a fine of up to 300 000 Rubles (USD4 000) and 15 days in jail for repeat offences."

[22] The Law Library of Congress, 'Initiatives to Counter Fake News in Selected Countries' (2019) (accessible at https://www.loc.gov/law/help/fake-news/counter-fake-news.pdf).

- **Kenya:** Kenya's Computer Misuse and Cybercrime Act criminalises the 'publication of false information in print, broadcast, data or over a computer system' in Articles 22 and 23. Despite legal challenges to various provisions that were alleged to stifle freedom of expression online, the Act was upheld as constitutional and came into effect in 2020.

- **COVID-19 false news laws:** the COVID-19 pandemic sparked a raft of oppressive false news laws across the world. The Disinformation Tracker, a collaborative civil society initiative, has documented the various responses, including laws criminalising false publications, initiated across the continent.

The criminalisation of the dissemination of fake news is likely to increase and may cause significant violence to freedom of expression. Such developments should be closely monitored and challenged where necessary. Fortunately, criminalisation is not the only option in addressing the rise of disinformation. Media and information literacy campaigns can effectively counter disinformation but providing a flood of accurate, reliable information instead and immunising audiences against false information before they are exposed to it. International bodies, states and CSOs are continually presenting new and innovative ways to address disinformation. Some notable contributions from international bodies include:

- **UNESCO**: UNESCO has developed the Journalism, fake news & disinformation: Handbook for journalism education and training. The handbook shares international good practices and serves as an internationally-relevant model curriculum, open to adoption or adaptation, which responds to the emerging global problem of disinformation that confronts societies in general, and journalism in particular.

- **European Union**: In 2018, the European Union published its Code of Practice on Disinformation. The purpose of the Code is to identify the actions that signatories could put in place in order to address the challenges related to disinformation. The Code discusses the need for safeguards against disinformation, implementation of reasonable policies, effective measures to close discernible fake accounts; and the improvement of the scrutiny of advertisement placements. The Code identifies best practices that signatories – such as Facebook, Google, Twitter, and Mozilla – should apply when implementing the Code's commitments.

- **Viral Facts Africa**: In response to the flood of COVID-19 mis- and disinformation on social media, the World Health Organisation (**WHO**) launched the Viral Facts Africa initiative, a network of fourteen fact-checking organisations and public health bodies that undertook health fact checks, explainers, myth busters and misinformation literacy messages optimised for sharing on Facebook, Twitter and Instagram. The initiative aims to rapidly debunk myths where they occur and provide viral, credible information.

At a state level, there have also been promising developments. In 2019, the US Library of Congress produced a report on Initiatives to Counter Fake News in Selected Countries. Some positive initiatives include:

- **Argentina:** The Commission for the Verification of Fake News was established. The Commission is envisaged to form part of the National Election Chamber, to assist with overcoming issues of disinformation during elections.

- **Sweden**: Bamse the Bear, a popular cartoon character in Sweden, has adopted a new role in teaching children about the dangers of disinformation by illustrating what happens to the bear's super-strength when false rumours are circulated about him.

- **Kenya:** The United States Embassy in Kenya started a media literacy campaign known as "YALI Checks: Stop.Reflect.Verify" to counter the spread of false information in Kenya. The campaign relies on an email series, an online quiz, blog posts, online chats, public outreach, educational videos, and an online pledge to engage with the Kenya chapter of the Young African Leaders Initiative (YALI) about disinformation.

- **Finland**: Finland has been lauded for winning the war on disinformation due to its initiatives aimed at teaching residents, students, journalists and politicians how to counter false information. The initiatives include courses at community colleges and the introduction of lessons in schools about disinformation.

- **Canada**: In January 2019, the Canadian government announced a multi-pronged effort to combat misinformation ahead of elections in the fall, comprised of four prongs. First, it created a "Critical Election Incident Public Protocol" to monitor and notify other agencies and the public about disinformation attempts, led by non-political officials. The government also called on social media platforms to do more to combat disinformation ahead of the election, in tandem with seeking to pass legislation to compel tech companies to be more transparent about their anti-disinformation and advertising policies. Third, Canada announced it was giving $7 million to projects aimed at increasing public awareness of misinformation online, and finally, the country launched a digital charter that set out principles for protecting freedom of expression while defending against online threats and disinformation.

Suggested standards for addressing disinformation

In the Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda, the following standards are suggested:

- General prohibitions on the dissemination of information based on vague and ambiguous ideas, including 'false news' or 'non-objective information', are incompatible with international standards for restrictions on freedom of expression, and should be abolished.
- Criminal defamation laws are unduly restrictive and should be abolished. Civil law rules on liability for false and defamatory statements are legitimate only if defendants are given a full opportunity and fail to prove the truth of those statements and also benefit from other defences, such as fair comment.

- State actors should not make, sponsor, encourage or further disseminate statements that they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda).
- State actors should, in accordance with their domestic and international legal obligations and their public duties, take care to ensure that they disseminate reliable and trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.

In line with these standards, the ACHPR's 2019 Declaration of Principles on Freedom of Expression and Access to Information in Africa provides under Principle 22 that States should repeal all laws that criminalise the publication of false news.

Determining limitations on freedom of expression

Global Partners Digital, in an attempt to determine how to tackle disinformation in a way that respects human rights, proposes an information-gathering approach to determine if disinformation amounts to a justifiable limitation of freedom of expression. Some of the suggested questions include:

- Is the basis for any restrictions on what information individuals can search for, receive, or impart set out in law?
- Is there clarity over the precise scope of the law so that individuals will know what is and is not restricted?
- Is speech restricted only where it is in pursuance of a legitimate aim?
- Are there exceptions or defences where the individual reasonably believed the information to be true?
- Are determinations made by an independent and impartial judicial authority?
- Are responses or sanctions proportionate?
- Is disinformation clearly defined?
- Are intermediaries liable for third-party content?
-

Fake news in the courts

In Africa, fake news laws have been challenged in the courts both domestically and at the regional level. In the case of *Chipenzi v The People* (2014), the High Court of Zambia found that a provision of Zambia's Penal Code that prohibited the publication of false information likely to cause public fear violated the Constitution as it did not amount to a reasonable justification for limiting the freedom of expression.

the Court of Justice of the Economic Community of West African States (**ECOWAS Court**) and the East African Court of Justice (**EACJ**) have both delivered landmark rulings on cases relating to the criminalisation of fake news.

In 2018, the ECOWAS Court decided the *Federation of African Journalists and Others v The Republic of The Gambia* matter, in which it considered offences of sedition, false news and criminal defamation in The Gambia's Criminal Code. Several journalists were arrested on charges of spreading false news. They argued that their rights to freedom of expression had been violated and sought a declaration from the Court that certain provisions of The Gambia's Criminal Code were inconsistent with regional and international law. The ECOWAS Court found that the criminal laws of the Gambia imposed criminal sanctions that are disproportionate and not necessary in a democratic society where freedom of speech is a guaranteed right and ordered that the legislation be reviewed. The Criminal Code was found to be broad and capable of casting an:

> "[E]xcessive burden upon the applicants in particular and all those who would exercise their right of free speech and violates the enshrined rights to freedom of speech and expression under Article 9 of the African Charter, Article 19 of the ICCPR and Article 19 of UDHR".

More recent developments in respect of the criminalisation of fake news came from the EACJ in the matter between the *Media Council of Tanzania and Others v Attorney-General of the United Republic of Tanzania*. In this case, the applicants challenged various provisions of the Tanzanian Media Services Act on the basis that "the Act in its current form is an unjustified restriction on the freedom of expression which is a cornerstone of the principles of democracy, the rule of law, accountability, transparency and good governance which [Tanzania] has committed to abide by, through the Treaty." The applicants argued that it violated freedom of expression by restricting the types of news or content without reasonable justification, criminalising the publication of false news and rumours, criminalising seditious statements, and vesting the Minister with absolute power to prohibit the import of publications or to sanction media content. The respondent argued that all the provisions are just and did not violate the right to freedom of expression and associated rights.

The EACJ held that although the sections were set out in law, the contents of these sections were vague, unclear, and imprecise. It noted that the use of the word "undermine" in the impugned provision, which formed the basis of the offence, was too vague to provide assurance to a journalist or other person who sought to regulate their conduct within the law. The EACJ further noted that the words "impede", "hate speech", "unwanted invasion", "infringe lawful commercial interests", "hinder or cause substantial harm", "significantly undermines" and "damage the information holder's position" are too broad or vague.

It further stated that it was persuaded by the applicants' submissions that section 52(1) of the Act failed the test of clarity and certainty. In this regard, it noted that definitions of sedition hinged on the possible and potential subjective reactions of audiences to whom the publication was made. This makes it impossible for a journalist or other individual to predict and thus plan their actions. In conclusion, the EACJ found in favour of the applicants and declared that, among other things, all the challenged provisions were in violation of articles 6(d) and 7(2) of the Treaty for the establishment of the East African Court of Justice (**EACJ Treaty**) and directed the Republic of Tanzania to take such measures as are necessary to bring the Media Services Act in compliance with the EACJ Treaty.

In an interesting case that addressed disinformation on social media, the High Court in South Africa in 2019 awarded damages to a public official who had been subject to a defamatory statement made by an opposition political party accusing him of nepotism and corruption. In the case of *Manuel v Economic Freedom Fighters and Others*, the court found that the political party had failed to prove the statement was true and taken no steps to verify its truthfulness, had published the tweet unreasonably, and had acted "with reckless indifference as to whether it was true or false". Most notably, the court held that the reasonable publication defence is not only available to the media:

> "Because of social media platforms like Twitter, Facebook and others, ordinary members of society now have publishing capacities capable of reaching beyond that which the print and broadcast media can".

On appeal, the damages award was subsequently overturned while the finding of defamation was upheld.

These landmark judgments provide guidance on the appropriate balance between legislating disinformation and protecting freedom of expression, and it is hoped they will have a far-reaching impact on other jurisdictions across the African region in ensuring that any responses to disinformation are based on international freedom of expression standards.

Defamation

Defamation is an important legal remedy for people whose reputation and dignity are harmed by the statements or actions of others. However, it is also frequently abused to unjustly stifle dissent. In particular, criminalising defamation is generally considered, under international human rights law, to be disproportionate and an unjustifiable infringement on the right to freedom of expression. The spread of the internet, and particularly social media platforms, has made it easier than ever to publish content to a wide audience, resulting in a rise in defamation being used against critical statements published online, and in speech that should be protected being criminalised under criminal defamation laws.

Overview of international instruments

The foundation for defamation in international law is article 17 of the ICCPR, which provides for protection against unlawful attacks on a person's honour and reputation. Article 19(3) of the ICCPR also refers to the rights and reputation of others as a legitimate ground for limiting the right to freedom of expression.[23] Reputation is therefore the underlying basis in any claim of defamation, whether slander or libel.[24]

---

[23] ICCPR: International Covenant on Civil and Political Rights (1976) (accessible at https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx).

[24] For a fuller discussion on the law on defamation, see the training manual published by Media Defence on the principles of freedom of expression under international law: Richard Carver, 'Training manual on international and comparative media and freedom of expression law', Media Defence at pp 48-64 (2018) (accessible at: https://www.mediadefence.org/sites/default/files/resources/files/MLDI.FoEManual.Version1.1.pdf). See also above no. 6 for a definition of libel and slander.

It is also noteworthy that in 2010 the ACHPR issued a Resolution calling on states to repeal criminal defamation laws or insult laws.[25]

Defamation in the courts

In recent years, many countries around the world have taken steps to decriminalise defamation in line with human rights standards. The UN Human Rights Council (UNHRC) General Comment No. 34 provides that: "States Parties should consider the decriminalisation of defamation and, in any case, the application of the criminal law should only be countenanced in the most serious of cases and imprisonment is never an appropriate penalty".[26] Principle 22 of the Declaration of Principles on Freedom of Expression and Access to Information in Africa calls on States to amend criminal laws on defamation and libel in favour of civil sanctions, and that the imposition of custodial sentences for defamation is a violation of the right to freedom of expression.

Several recent judgments across Africa demonstrate this trend, including the 2013 matter of *Konaté v Burkina Faso* in the African Court on Human and Peoples' Rights, *Misa-Zimbabwe et al v Minister of Justice et al in the Zimbabwe Constitutional Court*, *Peta v Minister of Law, Constitutional Affairs and Human Rights* in the Constitutional Court of Lesotho, and the 2018 case of *Federation of African Journalists and Others v The Gambia* in the ECOWAS Court. Most recently, the ACHPR ruled that Rwanda's criminal defamation laws violated freedom of expression and impeded development in democracies. It noted that such laws "constitute a serious interference with freedom of expression, impeding the public's right to access information, and the role of the media as a watchdog, preventing journalists and media practitioners from practising their profession in good faith, without fear of censorship".

Despite this, some countries, including South Africa and Zambia retain criminal defamation laws, underscoring the need for advocacy and litigation to address the situation.

The growth of Strategic Lawsuits Against Public Participation (SLAPP) suits by corporate actors using defamation laws to silence or intimidate is another concerning contemporary development that needs to be challenged. The ECtHR referred for the first time to the notion of a SLAPP suit in *OOO Memo v Russia* (2022) which involved a civil defamation suit brought by a Russian regional state body against a media company. In *Koko v Tanton* (2021), the Johannesburg High Court in South Africa held that a defamation case brought by a former executive of a state entity constituted a SLAPP suit.

Conclusion

The criminalisation of online speech presents an affront to the exercise of the right to freedom of expression online. However, as illustrated above, there are competing interests that need to be considered. With the rise of nefarious activities and feeble excuses from governments, it is important, now more than ever, that activists, lawyers, and individuals ensure that freedom

[25] ACHPR/Res.169(XLVIII)10, 'Resolution on Repealing Criminal Defamation Laws in Africa,' (2010) (accessible at: https://www.achpr.org/sessions/resolutions?id=343).
[26] UN Human Rights Council, 'General Comment No. 34 at article 47 (2011) (accessible at https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf).

of expression is protected, and only limited in terms of the clear prescripts of international human rights law.