Strathmore University
Centre for Intellectual Property and Information Technology Law

Internews
Local voices. Global change.

SERIES ON DIGITAL RIGHTS AND INTERNET FREEDOM

Topic 2: Freedom of Expression Online

Greater Internet Freedom

Centre for Intellectual Property and Information Technology Law (CIPIT)
Strathmore University

August 2023

Freedom of Expression Online

Author: The Centre for Intellectual Property and Information Technology Law (CIPIT).

Acknowledgements: We would like to express our gratitude to the Centre for Intellectual Property and Information Technology Law (CIPIT) acknowledging, Florence Ogonjo, Joshua Kitili, Lilian Olivia Orero, Doreen Aoko Abiero, Josephine Kaaniru, and Dan Allan Kipkoech who prepared this learning material. The CIPIT team authored this material in close consultation with the Greater Internet Freedom team at Internews, including Sigi Waigumo Mwanzia, Digital Rights Advisor, and Olga Kyryliuk, Technical Advisor on Internet Governance and Digital Rights.

Copy-Edited by: Internews.

Design & Layout by: CIPIT.

Strathmore University
Centre for Intellectual Property and Information Technology Law

Internews
Local voices. Global change.

About CIPIT	About GIF
The Centre for Intellectual Property and Information Technology Law (CIPIT) is an evidence-based research and training Centre based at Strathmore University, Nairobi, Kenya. CIPIT was established in 2012 and focuses on studying, creating, and sharing knowledge on the development of intellectual property and information technology utilizing diverse methodological approaches to inform debates on ICT applications and regulation.	**The Greater Internet Freedom Project (GIF) is a three-year, consortium-based, global program implemented by Internews and the GIF consortium across 39 countries. GIF places regional and local organizations at the forefront of the fight to preserve an open, reliable, secure, and interoperable Internet – and, by extension, protects the citizens, civic actors, journalists, and human rights defenders who rely on it to realize fundamental freedoms.**

Table of Contents

Introduction

The CIPIT and the GIF have developed exploratory material relevant to pertinent digital rights and internet freedom topics. The 'Freedom of Expression' topic examines efforts to promote the unhindered exercise of freedom of expression online. It delves into three issues impacting freedom of expression online, including legal restrictions, hate speech and information disorders, and online abuse, including cyberbullying and online harassment.

Online freedom of expression (FoE) acts as a cornerstone of democracy, fostering open dialogue, ensuring government accountability, and promoting individual liberties. FoE online empowers individuals to challenge authority, hold governments accountable, and advocate for social change.

The ubiquity of digital technologies and platforms, such as the internet, social media platforms and online forums, has played a critical role in enabling and facilitating FoE online worldwide, providing individuals with unprecedented access to information and diverse viewpoints. These technologies and platforms have further provided a global platform that facilitates social, cultural, economic, and political activism. They have enabled citizens to exercise their right to free speech, leading to increased participation in political discourse. This participation, in turn, can stimulate more robust democratic processes, and contribute to the creation of more balanced and just societies.

The digital environment, while opening new avenues for communication and expression, also presents new, complex challenges to the protection and promotion of FoE online. The 2011 uprisings in North Africa and the Middle East, often referred to as the Arab Spring, showcased the internet's potential as a platform for FoE and activism. Social media platforms became the key conduits for sharing information, organizing protests, and mobilizing citizens. However, these events also led governments to

perceive the internet as a potential threat, leading to the enactment of more stringent controls on online content.

The hopeful narrative of the internet as a liberating force came along with stricter state control, highlighting the realities of online FoE in the contemporary world. Balancing the need for open dialogue with the obligation to safeguard individual rights and reputations, and protect national security, public order, public health, or morals, remains a formidable challenge for societies.

Resource: ARTICLE 19's Global Expression Report, 2023

ARTICLE 19, an international freedom of expression NGO, notes that "freedom of expression is under threat and in decline…These threats are posed not just by autocratic governments, but also by legislation and law enforcement within democratic structures that erode the enjoyment of human rights, as well as by corporate interests and organised crime. Where those groups and their interests overlap, freedoms are in acute danger."

Source: ARTICLE 19.

This paper restricts itself to an exploration of the following complex challenges:

Legal Restrictions on the Right to Free Speech: The global nature of the internet poses legal challenges. Laws governing online speech vary greatly across different countries, and what is considered legal in one jurisdiction might not be in another.

- *This legal ambiguity continues to be exploited to limit FoE online.*

Online Censorship: Governments, corporations, and individuals continue to exploit their power to restrict, manipulate and censor online discourse, across all levels of the "internet stack," and at the technical, policy, and legal layers.

- *Traditional forms of governmental control and censorship persist. Online platforms, under the guise of moderating content, are excessively censoring and suppressing individuals' voices, thereby hindering free expression. Further, the deployment of new technologies, such as artificial intelligence (AI), for automated content moderation introduced new complexities.*

Hate Speech and Information Disorders: The proliferation of hate speech and online information disorders encompassing malinformation, misinformation, and disinformation, remains a central challenge and negatively impacts FoE online. With the advent of the internet, the dissemination of such speech has become alarmingly quick and widespread. Further, distinguishing between credible sources and false narratives can be difficult.

- *Hate speech can incite violence, discrimination, and other harmful behaviors, thus curtailing FoE for the targeted individuals or groups. In a similar vein, information disorders, undermine the integrity of online dialogue, and complicate efforts to engage in meaningful, fact-based discussions online.*

Cyberbullying and Online Harassment: These two prevalent risks are deterring individuals from freely expressing their thoughts and opinions online. The ubiquity of deepfakes – digital artifacts that can be used to manipulate images, audio recordings, and videos – is another threat to freedom of expression online.

- *Deepfakes have been used to spread disinformation or discredit vulnerable individuals, impacting political discourse, and turning online users into targets of online exploitation, further silencing groups.*

Definitions and Explainers

Term	Definition/Explainer
Digital/Internet/Online Censorship	The control or suppression of what can be accessed, published, or viewed on the Internet.[1]
Information Disorders	A term that details the pervasive nature and complexities of the information pollution phenomenon. It differentiates between mis-, dis- and mal-information, based on distinctions of harm and falseness.[2]
Internet Freedom	The exercise of internationally recognized human rights online… including the freedom to seek or impart information and ideas of all kinds regardless of frontiers through any medium.[3]

Freedom of Expression Online: International Instruments

The right to FoE, offline and online, is a foundational human right that upholds the democratic process in the physical and digital realms. This is also an enabling right that facilitates the enjoyment and fulfilment of other collective human rights online, by enabling individuals to freely express their opinions, ideas, convictions, beliefs, and creativity. Critically, it is important for individuals to be able to express their views on any number of issues, including those that are considered offensive or distasteful, in an enabling environment.

The right to FoE is enshrined in numerous international human rights treaties (*see table 1*) and in national constitutions.[4] International human rights law obligates states to protect and promote human rights, while businesses have a responsibility to respect human rights.[5]

The UN Human Rights Council (UNHRC) affirmed that the 'same rights that people have offline must also be protected online.'[6] Globally, the right to FoE is guaranteed and protected under Article 19 of the Universal Declaration of Human Rights (UDHR), and legal force is granted through Article 19 of the International Covenant on Civil and Political Rights (ICCPR).

The right to FoE online is **not absolute**. Under international human rights law, there are exceptional circumstances where the right to FoE online may be limited by States. However, these restrictions must satisfy the three-part test on legality, legitimacy, and necessity.[7] The Human Rights Committee emphasized that restrictions on electronic and online communication must meet the same criteria as offline communication.[8]

Resources: Three-Part Test on Permissible Restrictions of FoE Online

The ICCPR details permissible restrictions on online freedom of expression under Article 19 (2) (*the three-part test*) and Article 20, which prohibits propaganda for war and the advocacy of hatred. Critically, these restrictions "*must be narrowly interpreted and the necessity for any restrictions must be convincingly established."*[9]

Permissible Limitations – Three-Part Test:

- **Provided for by law:** any law or regulation must be formulated with sufficient precision to enable individuals to regulate their conduct accordingly.
- **In pursuit of a legitimate aim:** listed exhaustively as the respect of the rights or reputations of others, or the protection of national security or public order (ordre public), or of public health or morals.
- **Necessary and proportionate in a democratic society:** if a less intrusive measure can achieve the same purpose as a more restrictive one, the less restrictive measure must be applied. The UN HRC, in General Comment 34, reiterated this point, noting that restrictions of the right to FoE be proportionate and should not jeopardize the right itself: "*the relation between right and restriction and between norm and exception must not be reversed.*"[10]

Prohibition of War Propaganda and Advocacy of Hatred

Article 20 (1) of the ICCPR prohibits war propaganda. Article 20 (2) of the ICCPR, prohibits any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. This applies to communication or expression shared over the internet.

Sources: ICCPR. The European Court of Human Rights.

In the Africa region, the protection of online FoE is primarily guaranteed by the African Charter on Human and Peoples' Rights, specifically Article 9, which asserts 'every individual's right to receive information and express their opinions freely.'[11] Soft law provisions are outlined in the Declaration of Principles of Freedom of Expression and Access to Information in Africa (the Declaration). Principle 5 of the Declaration states that the "exercise of the rights to freedom of expression and access to information shall be protected from interference both online and offline."[12]

Further, the African Declaration on Internet Rights and Freedoms (ADIRF), developed by a Pan-African consortium, sets out 13 principles to uphold human and people's rights on the internet, including FoE. Principle 3 on Freedom of Expression affirms that

“everyone has the right to hold opinions without interference” and that the right to FoE includes “freedom to seek, receive and impart information and ideas of all kinds through the Internet and digital technologies and regardless of frontiers.”[13] The African Court on Human and Peoples' Rights, the continental court established to ensure the protection of human and peoples' rights, plays a significant role in enforcing these principles. The protection and promotion of FoE online varies across African countries, and restrictive laws and policies pose a significant challenge to the realization of these rights.

In the Central Asia (CA) region, the right to FoE online is recognized under Article 19 of the UDHR, which is adopted by many CA countries. However, the practical application of this right is often impeded by national laws and regulations that limit internet freedom.[14]

Despite these challenges, there are efforts at both the regional and national level to enhance online FoE. For instance, the Organization for Security and Co-operation in Europe (OSCE) works with member states in Central Asia to promote media freedom and support digital literacy initiatives. Further, the UN’s Human Rights Office of the High Commissioner for Human Rights' (OHCHR) established a Regional Office for Central Asia (ROCA) in 2008. The ROCA engages stakeholders to promote and protect human rights in Central Asia and strengthen compliance and protection against violations in all five countries in the CA region, including Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan.[15]

In the Balkan region, the European Convention on Human Rights, specifically Article 10, provides a basis for the protection of FoE online.[16] This right is also reinforced by Article 11 of the European Union’s (EU) Charter of Fundamental Rights, which reiterates that this right must be held without interference by public authorities.

However, the application of this right varies across nations within the Balkans, and in some cases, national laws are used to curtail online freedoms.

Turning to the Latin America and the Caribbean region, the Inter-American Court of Human Rights plays a vital role in protecting online FoE. The American Declaration of the Rights and Duties of Man and the American Convention on Human Rights both affirm the right to freedom of thought and expression, including the freedom to seek, receive, and impart information and ideas of all kinds.

Lastly, in the South and Southeast (SSE) Asia region, the right to FoE online is recognized under Article 19 of the Universal Declaration of Human Rights, which all countries in the region have adopted. Principle 23 of the Association of Southeast Asian Nations (ASEAN) Human Rights Declaration provides for the right to freedom of opinion and expression. However, restrictive national laws, internet shutdowns, and online censorship pose significant challenges to the realization of these rights. The ASEAN continues to take steps to address these issues and promote digital rights, but progress has been uneven across the region.

Table 1: Selected Resources - Freedom of Expression, International and Regional Instruments

Global Instruments
Universal Declaration of Human Rights (UDHR), 1948
International Covenant on Civil and Political Rights (ICCPR), 1966
Africa: Regional Instruments
African Charter on Human and Peoples' Rights (ACHPR), 1986
Declaration of Principles on Freedom of Expression in Africa, revised in 2019
Guidelines on Freedom of Association and Assembly in Africa, 2017

Asia: Regional Instruments
Association of Southeast Asian Nations (ASEAN) Human Rights Declaration, 2009
Balkans (Europe): Regional Instruments
Charter of Fundamental Rights of the European Union, 2009
Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, or ECHR), 1953
Latin America and the Caribbean: Regional Instruments
American Declaration of the Rights and Duties of Man (1948)
American Convention on Human Rights (1969)
Inter-American Democratic Charter (2001)

Legal Restrictions on Freedom of Expression Online

The most significant threat to FoE online is the adoption of broad, illegitimate, and disproportionate civil and criminal laws and policy instruments by states. Across different regions, when used to restrict FoE online, civil, and criminal laws generally aim to protect entities' (individuals, companies, the state etc.,) rights and reputations and address concerns related to national security, public order, or moral values, amongst others. Further, digital providers, such as social media platforms, internet intermediaries, hosting platforms, are required to restrict FoE online, as a means of complying with national laws, or through their internal policy instruments, such as terms of service.

Notably, any direct or indirect measures taken by States to limit FoE amount to a "*prima facie interference with the right*."[17]

Table 2: Common Laws Impacting FoE Online

Laws Impacting FoE Online Given jurisdictional differences, the type and nature of laws adopted by governments differ radically. These differences are informed by (a) the type of government/political system in place (i.e., democratic, authoritarian etc), (b) the socio-economic conditions in the country, (c) the government's tolerance levels for criticism and dissent, (d) the levels of internet penetration, digital adoption, and innovation, (e) cultural attitudes towards freedom of expression, amongst others. The laws detailed below are not exhaustive, and often contain a mixture of both civil and criminal provisions. Notably, the mere existence of laws can have a 'chilling effect' on FoE online, leading individuals and communities to self-censor their speech, communication, and activities online due to the fear of censure.	
Civil Laws	**Criminal Laws**
• **Civil defamation laws:** Prohibiting the dissemination of online defamatory statements that harm a person's reputation. • **Copyright/intellectual property (IP) laws:**	• **Blasphemy laws:** Criminalizing speech considered disrespectful or offensive towards religious beliefs or symbols. • **Cybercrime/computer laws:**

Regulating the use and distribution of copyrighted materials and IP rights related to inventions, trademarks, and creative works. The balance between IP/copyright holders' rights with the public's interest can clash where material is used for commentary, criticism, parody. This content, if used without authorisation, can result in takedowns, content removal, or legal action. • **Cyberbullying laws:** Regulating specific behaviors such as persistent online bullying, harassment, threats, stalking, amongst others. These laws differentiate between protected speech and harmful behaviour. • **Cybersecurity laws:** Enforcing measures to protect online systems, networks, and data. These laws permit content filtering or blocking to address cybersecurity threats. • **Digital taxation laws:** Regulating the taxation of digital services or online transactions. These laws have the potential for unequal impact creating an economic burden on e.g., indigent populations, thereby preventing their ability to exercise their FoE online. • **Electronic surveillance laws:** Regulating the monitoring and interception of electronic communications for security or law enforcement purposes. These laws create a chilling effect on FoE online due to the fear of being monitored or surveilled. • **False/"fake" news laws:** Regulating the dissemination of false or misleading information online. • **Hate speech laws:** Restricting speech that incites violence or promotes discrimination based on characteristics such as race,	Addressing online activities such as offensive communication, fake news, cyberbullying, online harassment, hacking, identity theft, or computer fraud. • **Criminal defamation laws:** Criminalisation of the dissemination of online defamatory statement that harms a person's reputation. • **Harassment laws:** Addressing persistent and unwanted online behaviour that causes distress or fear to individuals. • **Incitement to violence laws:** Prohibiting speech that directly encourages or promotes violent acts. • **National security laws:** Restricting speech that poses a threat to national security, including dissemination of classified information. • **Obscenity laws:** Criminalizing the dissemination of sexually explicit or pornographic materials. • **Penal & Criminal Codes:** Regulating a broad array of speech and content related offences. These generally enable surveillance and monitoring activities and grant wide powers to law enforcement agencies to investigate and prosecute FoE-related offenses. • **Public health laws:** Regulating misinformation or harmful content that risks public health

religion, or ethnicity. • **Online content regulation laws:** Regulating the content published online, including age restrictions, obscenity, or harmful material. • **Privacy and data protection laws:** Regulating the unauthorized disclosure of personal or confidential information and the collection, storage, and processing of personal data online. • **Right to be forgotten laws:** Allowing individuals to request removal of certain online information about themselves. • **Telecommunications/ICT laws:** Regulating the use and operation of telecommunication and information and communication technology services. Often contain provisions on online content moderation, internet intermediary liability provisions, amongst others that require telecoms provides to regulate online access and content.	efforts. These were used in varying degrees during the COVID-19 pandemic. • **Public order laws:** Regulating speech that disrupts public peace or incites public unrest. • **Terrorism laws:** Targeting speech or expression related to terrorist activities or promoting violence. • **Sedition laws:** Criminalizing speech or expression that incites rebellion or promotes the overthrow of the government. • **State secrets laws:** Criminalizing the disclosure of classified or sensitive government information.
Sources: Colombia University. GNI. UNESCO. ARTICLE 19. CIMA.	

While narrowly interpreted and necessary restrictions provided under international law on FoE online are permissible, states globally are enacting ***overly broad*** laws, policies, and regulations that risk rendering the international protection of this right meaningless. These laws spark concerns about governmental overreach and the stifling of criticism. They also present a challenging paradox - the need for laws to maintain order and protect individuals from harmful content online, versus the potential misuse of these laws to suppress dissent and limit FoE online.

The following types of permissible speech and communication continue to be regulated by states, often in contravention of international standards and laws:[18]

Offensive Speech and Communication: this permissible speech is heavily regulated by states worldwide, and is frequently legislated **under penal codes, defamation laws, computer misuse, and cybercrimes laws.** Critically, what one might view as an exercise of their right to free speech, another might perceive as a direct attack or insult. In the digital realm, this issue manifests as a tightrope walk between safeguarding individual rights to express opinions freely and maintaining an environment of respect and tolerance. Under international law, "expression that is offensive, disturbing or shocking"[19] is protected, unless it incites genocide, or advocates for discriminatory hatred that constitutes incitement to violence, hostility, or discrimination. [20]

Offensive Communication in Uganda

On January 10, 2023, Uganda's Constitutional Court struck down Section 25 of the Computer Misuse Act, 2011, which criminalised 'offensive communication,' on grounds that this contradicted the constitutional protection of freedom of speech and expression. This provision defined offensive communication as the 'wilful and repeated use of electronic communication to disturb or attempt to disturb the peace, quiet or right of privacy of any person with no purpose of legitimate communication whether or not a conversation ensues."

This provision was used to charge Stella Nyanzi, a Ugandan poet, who published a poem on Facebook on H.E. President Yoweri Museveni's birthday. The poem is an excellent example of the reach of FoE online, and the protection of what many viewed as deeply offensive and lewd language.

An excerpt of the poem follows:

"Yoweri, they say it was your birthday yesterday.
How bitterly sad a day!
I wish the smelly and itchy cream-colored candida festering in Esiteri's cunt had suffocated you to death during birth.
Suffocated you just like you are suffocating us with oppression, suppression, and repression!

Yoweri, they say it was your birthday yesterday.
How painfully ugly a day!
I wish the lice-filled bush of dirty pubic hair overgrown all over Esiteri's unwashed chuchu had strangled you at birth.
Strangled you just like the long tentacles of corruption you sowed and watered into our bleeding

economy…"

Sources: *Committee to Protect Journalists*. *NPR*. *Motivation Africa.*

Satire and Parody: these permissible forms of speech are regulated by some states, and are frequently restricted under **defamation, blasphemy, "fake-news", disinformation and *lese-majeste* laws.** The legislation of satire and parody under "fake news" and disinformation laws fails to recognize that "parody and satire do not easily fall into a binary analysis of truth and falsity."[21]

Satire and parody are critical forms of artistic expression and continue to be used as a form of expressing public opinion. Satire and parody are also used to critique societal norms, religions or belief systems, or political regimes. Satire is a form of humor that uses ridicule, irony, or sarcasm to criticize human vices or flaws. It often involves exaggerating or distorting the characteristics of a particular person, group, or institution for comedic effect.[22] Parody, on the other hand, is an imitation of another work with humorous intent. It imitates the style and content of other works to make them appear ridiculous. [23] The Internet has democratized content creation, enabling creatives can express their wit, critique, ridicule, and commentary without the traditional gatekeepers.

Under international human rights law, humor, satire, and parody content are all forms of protected speech, unless this incites genocide, or advocates for discriminatory hatred that constitutes incitement to violence, hostility, or discrimination. [24]In recent years, some states have attempted to criminalize the posting of satirical or parody content online, citing public order and morality to justify their actions.

Parody Charges in the United States of America

In 2016, Anthony Novak created a satirical Facebook account that parodied the Facebook account of the Ohio Police Department. Novak parodied the department's slogan, "*We know crime*" to read "*We no crime*." Novak was arrested and charged with the offence of using a computer to disrupt police operations. The police argued that Novak's account had made the leap from satire to posing an actual risk to public safety. Novak was acquitted by a jury in August 2016.

Sources: *The New York Times*. *Ohio Capital Journal*. *Cleveland.com*.

Religious Blasphemy or "defamation of religion": states worldwide are legislating conversations on religion and other belief systems relying on **anti-blasphemy laws** under **'**blasphemy,' 'defamation of religion,' or 'protecting religion or religious sensitivities' provisions. Generally, these laws restrict FoE online by prohibiting insults or the display of a lack of respect for a religion or a belief system. Unless these restrictions are narrowly framed per permissible restrictions under Articles 19 (3) and 20 (2) of the ICCPR, these are incompatible with international human rights law.[25] Recently, a number of European countries such as the United Kingdom, Denmark, and Canada repealed their blasphemy laws in 2008, 2017 and 2018 respectively.[26]

Religious Blasphemy Laws: Africa, Asia, and the Middle East

Ethiopia: under Articles 492 and 816 of the Criminal Code, public expressions (*including words, gestures, or scoffs*) mocking religion or deemed blasphemous, scandalous, or grossly offensive are punishable with a fine or up to one month of arrest. This expression can be directed towards individuals, the Divine Being, religious figures, authorised religious ceremonies or offices. It is also illegal to disrupt religious ceremonies or to desecrate religious places, images, or objects.[27]

Malaysia: various laws prohibit offenses against Islamic and any other religions including the Malaysian Criminal Code, the Penal Code, and the Communications Act. For example, it is a crime to utter words, make any sounds or a gesture or place an object with the deliberate intent of wounding a person's religious feelings.[28]

Saudi Arabia: Under the counterterrorism law, it is illegal to criticize Islamic beliefs or

practices. The law criminalises, *inter alia*:

- The "calling for atheist thought in any form or calling into question the fundamentals of the Islamic religion."
- "Anyone who challenges, either directly or indirectly, the religion or justice of the King or Crown Prince."[29]

Sources: *US Department of State*. *OHCHR*. *End Blasphemy Laws*.

Digital rights and Internet freedom organizations, coalitions and individual activists are actively working to challenge legal restrictions criminalizing online FoE, at both the policy and infrastructure levels.[30] Generally, these advocacy efforts fall under the following umbrellas: policy and legislative advocacy involving direct engagements with policymakers and legislators, public interest litigation, online and offline advocacy campaigns, amongst others. These efforts can take a long time to bear fruit, and are frequently subject to political will and interests, and accompanying societal changes.

Additionally, international experts, including human rights mechanisms, special rapporteurs dealing with human rights, amongst others, continue to collaborate on cross-regional issues related to FoE online. These efforts raise awareness about legal restrictions on online content at the regional and global levels.[31]

Online Censorship

Governments, corporations, and online communities continue to exploit their power to restrict, manipulate and censor online discourse at the technical, policy and legal levels. **Traditional forms of governmental control and censorship** at scale (i.e., at the national level), such as restricting or completely blocking access to websites, social media platforms, and online services and platforms, persist.

Conversely**, private actors**, such as social media platforms are censoring and suppressing individuals' voices as part of their content moderation efforts which hinders free expression at scale. The deployment of new technologies, such as artificial intelligence (AI) for automated content moderation, has introduced new complexities for the protection of FoE online. In addition to online censorship at the state and private sector levels, this also takes place at the institutional, regional, and community/individual levels.

Institutional-level censorship is frequently witnessed in workplaces, at schools and universities which implement censorship policies to regulate access to certain websites or types of content. These measures are usually intended to maintain productivity, protect against inappropriate or harmful material, or adhere to institutional guidelines. **Regional-level censorship** can involve blocking or restricting access to specific websites or platforms within certain regions based on regional laws, cultural sensitivities, or political considerations.

Further **community or individual-level censorship**, or **self-censorship involves** individuals and online communities practicing self-censorship by refraining from expressing certain views or sharing specific content due to concerns about repercussions, social norms, or fear of harassment. Self-censorship can be a result of the chilling effect caused by the presence of censorship measures.[32]

Generally, online or internet censorship is defined as the "control or suppression of what can be accessed, published, or viewed on the Internet."[33] This ongoing, global, practice is at the forefront of FoE debates because of its *negative, and scaled impact* on the ability of individuals to express themselves freely online, to access diverse content and viewpoints, and engage in open dialogue.

Despite internationally permissible grounds permitting the censorship of speech, information, and communication online on grounds of public order, national security, public health, or moral values, these must be narrowly construed, legitimate, and proportionately balanced with the preservation of FoE online.

Important Note

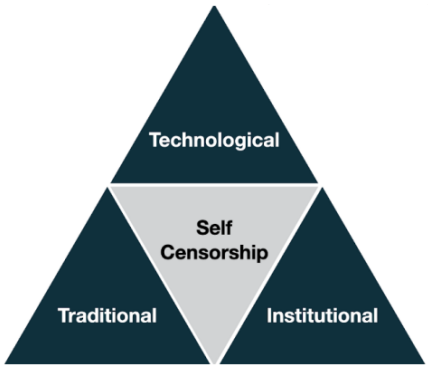
"Efforts to restrict access have developed in step with improved infrastructure and technology that should enable access. Technical measures are being implemented in many jurisdictions by state and non-state actors to limit, influence, monitor, and control people's access to the internet. These measures include censoring, blocking, filtering, and monitoring content. While these measures may not be as extreme as complete internet shutdowns, they equally hinder the full enjoyment of the right to freedom of expression and have the potential to severely distort and disrupt people's access to information online."

Source: Media Defence

State Actors

Online censorship by state actors is also referred to as digital censorship, which includes actions taken by a government to "remove or obscure internet content from its citizens or to limit the ability of someone to digitally transmit information to a broad audience."[34] Digital censorship activities are part of a larger online censorship ecosystem, where governments use a wide variety of 'pressure levers,' including laws, penalties, surveillance, and other methods, to restrict access to online platforms and content or push individuals to self-censor due to fear of punishment or harm (*see figure 1 below*).[35]

Figure 1: ***Components of a "Censorship Ecosystem" Fostering Self-Censorship***



Online censorship by both autocratic and democratic governments is at an all-time high around the world, is not a new phenomenon and challenges the long-held position that the Internet is a "technology that is difficult to censor."[36] The traditional censorship methods used by governments range from mandatory blocking at the infrastructure level (*e.g., entire websites (local and foreign), IP addresses*), content filtering (*a form of prior censorship*), online monitoring and the tracking and surveillance of users' online activities, amongst others.[37]

Important Note

Digital censorship by states can have a long-lasting and far-reaching effect on individuals' FoE and other human rights.

- It denies people their right to make informed decisions or form thoughtful opinions due to the limitation on individuals' ability to access certain types of information or resources.
- It can limit citizens' capacity to engage in healthy public debates or to challenge the status quo through, e.g., the restriction of digital media.
- It can lead to an overall decline in digital literacy and creativity, and a decrease in trust between citizens and their governments.
- It disempowers individuals, who are forced to use digital circumvention tools, to regain access to blocked, filtered, or restricted digital platforms.
- It disempowers people by making it difficult for individuals to hold governments accountable under the social contract through digital platforms.

Sources: European Parliament. OHCHR.

Given technological advancements, governments around the world are adopting traditional and non-traditional digital censorship measures, tools, and technologies to control what individuals access, see, and say online in their jurisdictions. The table below provides a general overview of digital censorship activities in Africa, the Balkans, Central Asia, Latin America and the Caribbean and South and Southeast Asia. Due to evolving political, social, technological, and legal circumstances, this table is subject to change.

Table 3: Digital Censorship by States

Region	Measures	Tools/Technologies
Africa	Content filtering and blocking, content removal requests, website blocking, internet shutdowns, social media monitoring and surveillance, Internet taxation, criminalization of online expression.	DNS tampering, deep packet inspection, keyword filtering, content monitoring tools, internet data centers, surveillance networks
Balkans	Website blocking, content removal requests, legal actions against online platforms, internet surveillance.	Internet traffic analysis techniques, DNS manipulation, software-based censorship technology such as firewalls.
Central Asia	Content restrictions, internet filtering, website blocking, surveillance and monitoring of online activities, criminalization of dissenting speech.	DNS tampering, network filtering, active probes and blocking requests to specific websites.
Latin America and the Caribbean	Website blocking, content removal requests, criminalization of online expression, surveillance and monitoring, legal action against social media platforms.	DNS manipulation, IP address blocking, packet filtering and traffic shaping technologies.
South and Southeast Asia	Content restrictions, website blocking, content filtering, social media monitoring, internet shutdowns, legal actions against online platforms.	IP address blocking, keyword filtering, DNS manipulation, web page blocking.
Sources: OONI. OpenNet Initiative. Freedom House. Massachusetts Institute of Technology.		

University of Berkeley.

Despite this, pockets of coordinated and sporadic resistance against digital censorship continue to spring up, even in countries with tight control. Digital rights and internet freedom organizations, with support from international human rights mechanisms around the world continue to:

a. Raise awareness about the deployment of detectable digital censorship activities by states before local, national, and international fora.
b. Empower users with the knowledge and tools to bypass digital censorship.[38]

Private Sector Actors

Private sector entities contribute to the online censorship ecosystem at the policymaking, infrastructure, and technology levels. This interferences with users' right to FoE online, and further impacts private sectors obligations under international human rights law to respect human rights. Generally, research and advocacy efforts have focused on three types of private sector entities in the 'tech stack' engaging in online censorship activities. These include social media platforms, and internet intermediaries, such as Internet Service Providers (ISPs) and telecommunications providers, whose online censorship activities have a large-scale impact on FoE online.[39]

Important Note: *The "Tech Stack"*

In 2022, the term 'tech stack' was used by 57 organizations to refer to a broader and wider set of private sector entities that impact FoE online.[40] These organizations called for the safeguarding of the 'stack,' specifically in the content moderation space.

> *"Users and policymakers are very familiar with platforms like Facebook, Twitter, or YouTube. But those services are not the internet. In fact, online communication and commerce also depend on a wide range of service providers, including ISPs and telcos, like Comcast, Orange, MTN, Airtel, Movistar, or Vodafone; domain name registrars such as Namecheap or GoDaddy; support services such as Amazon Web Services (AWS), certificate authorities (such as Let's Encrypt), payment processors such as PayPal and M-Pesa, email, messaging services, and more. Taken together, these providers are sometimes called the "tech stack."*

Sources: Protect the Stack. Washington Post.

This emphasis has often resulted in the **neglect of other private sector entities operating higher up the 'tech stack.'** These entities tend to receive less public scrutiny and their 'human rights obligations are not as widely recognized or understood, despite their significant roles and responsibilities in the online ecosystem. A few examples include the 'Internet Exchange Points (IXPs), domain registrars and registries, and technical standard setting bodies, such as the Internet Corporation for Assigned Names and Numbers (ICANN).'[41] Others include mobile network operators deploying technology standard for broadband cellular networks, such as 5G technology.

Content Moderation by Social Media Platforms

Important Note

In 2011, amidst the throes of the Egyptian Revolution, Asmaa Mahfouz championed the cause of FoE online. Through a video posted online, she sparked the outrage and conviction needed to mobilize the masses. Restricted by traditional societal norms and fear of government backlash, many had been hesitant to voice their discontentment. Mahfouz, leveraging the power of the internet, shattered those constraints. Her impassioned plea resonated, triggering a cascade of dissent that ultimately led to a transformative shift in power in Egypt. This anecdote embodies the tremendous potential of online platforms as tools for freedom of expression, paving the way for democratic dialogue and social change.

Source: Democracy Now

As illustrated by the Asmaa Mahfouz narrative above, **social media platforms** are powerful conduits for FoE online. They have democratized information dissemination, allowing individuals to create and share content on a global scale, without the need for intermediaries such as traditional media outlets, which had been subject to prior restraint efforts from governments.[42] Platforms like Facebook, Twitter, Threads, YouTube, and Instagram have become digital town squares, where individuals can voice their opinions, rally for causes, and connect with like-minded individuals.

However, these platforms are not without their complexities. They walk a delicate balance between respecting free expression online and dealing with 'harmful' or 'objectionable' content to comply with internal policies or standards or national or regional level laws or government directives.

Online Platforms and Intermediary Liability in the US

Section 230 of the Communications Decency Act of 1996 provides online platforms with intermediary liability protections, as follows: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."

Under the Trump administration, the passage of two laws, the Stop Enabling Sex Traffickers Act (SESTA) and the Fight Online Sex Trafficking Act (FOSTA), effectively suspended Section 230 by imposing civil and criminal liability on online platforms for "sexually explicit content that depicts either underage persons or non-consensual activity that appears on their sites."

Generally, such laws place online platforms in a position where they err on the side of caution, leading to the over censoring of all material that is deemed 'sexually explicit.' In 2018, the Electronic Frontier Foundation (EFF) filed a lawsuit challenging the law. EFF recognised that the law was enacted by the US Congress "for the worthy purpose of fighting sex trafficking" but decried the broad language used in the law as permitting the criminalisation of protected speech. Further, EFF notes that the "broad language makes criminals of those who advocate for and provide resources to adult, consensual sex workers and actually hinders efforts to prosecute sex traffickers and aid victims."

Source: Berkeley Journal of Criminal Law. EFF.

Platforms generally enforce **policies or community standards** against hate speech, misinformation, and online harassment to protect the rights and safety of all users. However, social media platforms have been accused of either (i) failing to apply their policies in a proportionate manner across the world leading to online and offline harms targeting specific groups,[43] or (ii) applying overly stringent controls that stifled free speech, leading to accusations of censorship by platforms.[44]

Social media platforms have immense power in shaping public discourse due to their control over content visibility through human and/or automated content moderation

activities. The **content moderation debate is integral to FoE online in the digital age,** with multiple stakeholders grappling to provide a workable solution to content moderation activities by platforms. The Trust and Safety Professional Association define content moderation as:

> *"the process of reviewing online user-generated content for compliance against a digital platform's policies regarding what is and what is not allowed to be shared on their platform. These policies are often known as community standards. The process of moderating content and enforcing policy is either done manually by people or through automation, or a combination of both, depending on the scale and maturity of the abuse and of a platform's operations... In practice, content moderation can mean moderating individual content and/or actors and their behavior on the platform."*[45]

The moderation of content by private sector entities has a significant impact on FoE online as it can effectively silence certain voices while leaving others alone. This can lead to a skewed representation of what is acceptable in an online space, with those voices that are most often unmoderated being the ones that hold sway. The silencing of voices online is compounded by the challenges of **automated moderation using algorithms and machine learning systems.**[46]

Automated moderation has given rise to **concerns of algorithmic censorship without human intervention**, leading to concerns about bias and the manipulation of online information.[47] Overall, transparency in content moderation policies and algorithmic processes is critical to ensure that social media platforms enable rather than restrict FoE online.[48]

As illustrated by Table 4 below, private sector companies, such as ISPs, telecommunications companies, search engines, and social media companies install and deploy systems and technologies that actively monitor and censor all types of content, amongst other online censorship activities. These measures, which range from content filtering to blocking access to certain websites or services, to suspending or deactivating users' accounts and the use of hardware and software censorship

technologies,[49] run the real risk of stifling legitimate and lawful speech, negatively impacting FoE online.

Given this negative impact, it is incumbent upon society, including users, policymakers, and the platforms themselves, to navigate these challenges within the remit of international human rights law, and ensure that the digital public square remains open and accessible to all.

Table 4: Online Censorship by Private Sector Actors

Entity	Measures/Tools	Impact on FoE Online
ISPs/Telecom Providers **(e.g., AT&T, Airtel)**	Blocking or throttling of services	Censorship of websites and content deemed offensive or illegal Limitation on net neutrality
	Content removal requests from governments or private companies	Suppression of information, opinions, and other forms of expression without a legal process
	Restrictions on providing services or hosting content	Increasingly limited access to online resources Restricted ability of service providers to host certain content
	Blocking of websites or platforms (*in full or in part*)	Limitations on access due to laws and regulations Severe restrictions on the ability of service providers to provide services or host content, leading to censorship without a legal process
	Internet filtering systems or traffic shaping techniques	Surveillance and monitoring of online users' activities, creating chilling effect on FoE online (self-censorship arising from fear of being monitored)

Search engines (e.g., Google, Bing)	Manual/automated content moderation	Removal or limitation of certain types of content, including those are that lawful, but controversial, or content is wrongly flagged and removed
	Content/keyword filtering	Inadvertent or deliberate blocking of lawful, but controversial, content, or key words
	Content rating	Suppressing of legitimate speech, impacting FoE
Social Media Companies (e.g., YouTube, Facebook, Twitter)	'Voluntary Codes' for Service Use (e.g., Terms of Service, and associated policies)	Allows companies to set their own rules for acceptable speech, sometimes going above and beyond international human rights law restrictions, and potentially limiting discourse
	Manual Content Detection/Removal	May unfairly target certain types of speech due to lack of contextual understanding by human moderators
	Automated Content Detection/Removal	May unfairly target certain types of speech due to biases in machine learning algorithms
	Content Filtering	Blocks certain words or phrases, limiting the ability to express some ideas or topics
	User Account Suspension, Disabling or Deactivation	Silences specific users, often without a clear or fair process for disputing the action, or with a complicated dispute resolution process
Sources: Meta. Penn Law. Media Defence.		

Hate Speech, Information Disorders and FoE Online

'Hate speech' and 'information disorders' are two terms that have become increasingly important in the digital age, given their negative impact on FoE online. While these are not new phenomena, digital and communications technologies have amplified their scale and impact on a global scale. The proliferation of hate speech on online platforms promotes the spread of divisive rhetoric which has the potential to cause offline harm, whereas information disorders, depending on the intent and harm, create distrust in the content that online users access online.

Hate Speech: There is no universally accepted definition of hate speech under international human rights law. The term is extremely emotive given its targeting of people, as individuals or groups, *because of who they are,* and the differing interpretations of what is and what isn't 'hateful.'

Important Note

Across the world, the term' hate speech' is conceptualised and applied differently.

United Nations: In 2019, the UN developed its Strategy and Plan of Action on Hate Speech, and adopted the following definition of hate speech:

> *"... any kind of communication in speech, writing or behavior that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."*

ARTICLE 19: In 2015, ARTICLE 19 stressed that the term 'hate speech' is a "broad concept that captures a wide range of expression" including protected speech. Based on this, the organisation advanced a three-tiered typology permitting the identification and differentiation of different forms of 'hate speech, based on severity:

1. *'Hate speech' that must be prohibited:*
2. *'Hate speech' that may be prohibited:*
3. *Lawful 'hate speech.'*

National Cohesion and Integration Commission (Kenya): The NCIC in Kenya adopts the following definition of hate speech, as defined under the National Cohesion and Integration Act, 2008:

"A person who—

a) uses threatening, abusive or insulting words or behaviour, or displays any written material;
b) publishes or distributes written material;
c) presents or directs the performance the public performance of a play;
d) distributes, shows or plays, a recording of visual images; or
e) provides, produces or directs a programme, which is threatening, abusive or insulting or involves the use of threatening, abusive or insulting words or behavior commits an offence if such person intends thereby to stir up ethnic hatred, or having regard to all the circumstances, ethnic hatred is likely to be stirred up."

Meta: Meta defines hate speech as

"...as a direct attack against people – rather than concepts or institutions – on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease."

Sources: UN. ARTICLE 19. European Parliament. Kenya Law. Meta.

Online hate speech can manifest itself in different forms. Concerningly, hate speech online has the potential to generate and fuel intolerance, hatred, divisiveness and offline violence and harm towards people. A few forms of hate speech include:

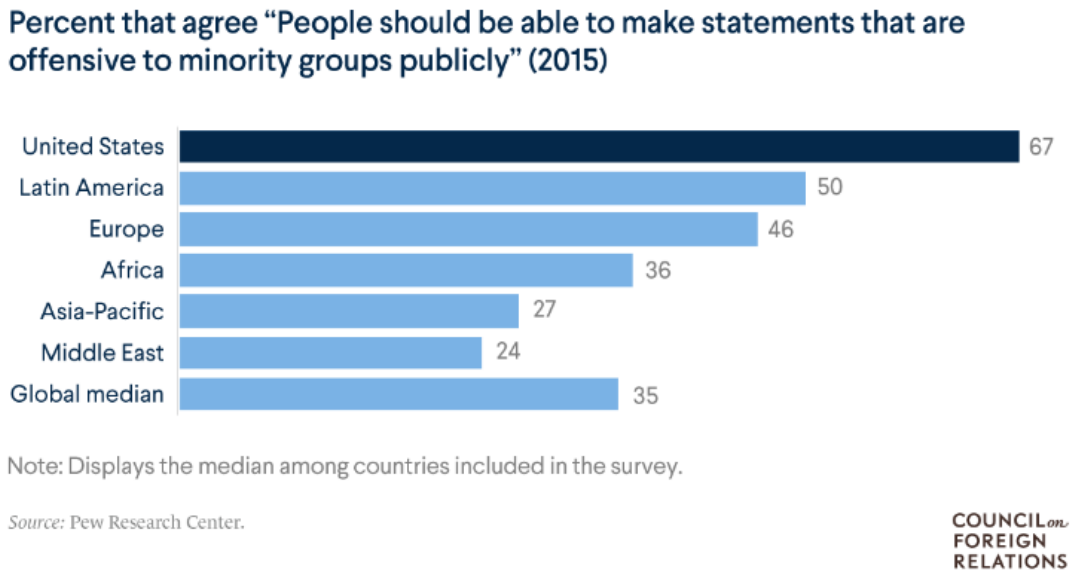
- Stirring hatred or violence using content (text, audio, images, videos, etc.) on online platforms, targeting protected characteristics, such as race, colour, sex, ethnicity, religion, gender identity, sexual orientation, migrant or refugee status, political opinions, amongst others.
- The creation of websites and online forums intended to spread hatred directed towards certain demographics.

As noted by the Special Rapporteur on Minority issues, "hate speech is followed by hate crimes and violence"[50] with online harm being traced from the online to the physical world.[51] In recent years, **hate speech has targeted minority groups, particularly national, ethnic, religious, sexual, and linguistic minorities.**[52] As illustrated by Figure 2 below, people residing in North America, Latin America and Europe believe that they should be able to issue offensive statements to minority groups

in public, compared to Africa, the Asia-Pacific, and the Middle East. In the aftermath of the New Zealand killing and injury of 51 worshippers at two mosques in Christchurch, reports traced the pervasive nature of hate speech worldwide, noting that:

"As more and more people have moved online, experts say, individuals inclined toward racism, misogyny, or homophobia have found niches that can reinforce their views and goad them to violence. Social media platforms also offer violent actors the opportunity to publicize their acts."[53]

Figure 2: ***Pew Research Centre***

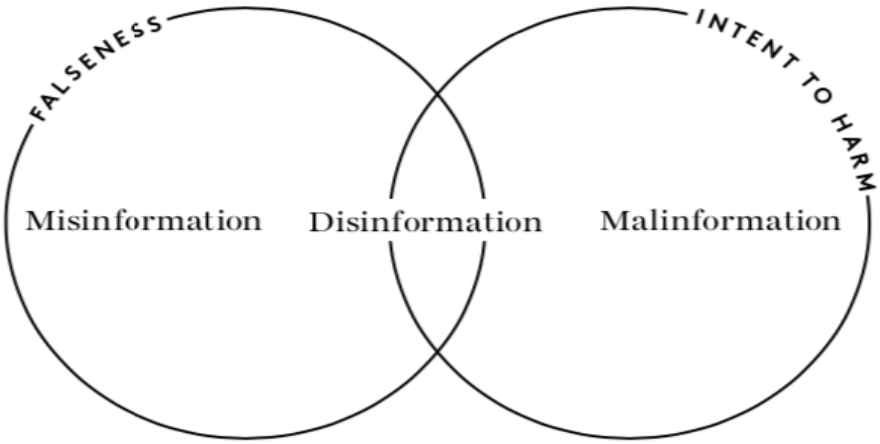


Information Disorders: This term – which encompasses malinformation, misinformation, and disinformation, was first used in 2017 by Dr Claire Wardle and Hossein Derakhshan.[54] These disorders impact the information ecosystem by introducing doubt about the authenticity and veracity of information online. As

illustrated by *Figure 3 below*, misinformation refers to the unintentional spread of false or inaccurate information, usually without any ill intent. Source confusion — lacking knowledge of where the content originated from — is often the cause for such accidental propagation.

Conversely, malicious disinformation (or malinformation) refers to the use of realistic information to intentionally deceive people and deliberately inflict harm on a person, organization or country. Disinformation straddles both extremes and includes the circulation of false information by a malicious actor with the deliberate intention of causing harm and deceiving people.

Figure 3: ***Understanding Information disorder. Source: First Draft.***



These forms of information can have a huge impact on FoE online, as they often go viral quickly and can be very hard to counter. Misinformation and disinformation are both used to spread lies or distort the truth so that people make decisions based on false information, which can in turn lead to censorship or threats against those who express unpopular opinions. Malinformation, on the other hand, can lead to harassment and discrimination of individuals by spreading their personal information or targeting them with false allegations.

To combat both hate speech and information disorders online, several organizations provide fact-checking and media literacy support to strengthen the integrity of the online information ecosystem. Additionally, several social media platforms have signed voluntary codes in the EU and Australia, to combat hate speech information disorders that are increasingly being spread through their platforms.

Important Note

Two voluntary self-regulation approaches have been taken in the content moderation space by online platforms and services. On one hand, an individual company opted to create a quasi-judicial entity to moderate its own moderation decisions, and on the other hand, self-regulation was coordinated through private-public cooperation.

- **Meta and the Oversight Board:**
  - In 2020, Meta (*formerly Facebook*) created the Oversight Board, an independent body with oversight over Meta's content moderation decisions on both Facebook and Instagram. The Oversight Board is tasked with promoting free expression on these two platforms, and its decisions to uphold or reverse Meta's moderation decisions are binding on Meta, unless these violate international law.
- **Voluntary Codes in the EU and Australia:**
  - **European Commission's 2022 Code of Practice on Disinformation:** To fight online disinformation in the EU, 44 signatories signed the Code of Practice and voluntarily agreed to internalise the Code into their relevant missions. Among the 44 signatories, there are online platforms and service providers (TikTok, Meta and Microsoft, ClubHouse, Google, ActiveFence, etc), players from the advertising ecosystem, fact-checkers, civil society, research, and other organisations.
  - **Voluntary Code of Practice on Misinformation and Disinformation:** Building on Australia's 2019 policy on Regulating in the Digital Age: Government Response and Implementation Roadmap, the Digital Industry Group Inc. (DIGI), a non-profit industry association developed the Australia Code of Practice on Disinformation and Misinformation, 2022. This code is overseen by the Australian Communications and Media Authority (ACMA). There are eight voluntary signatories, including Adobe, Apple, Google, Meta, Microsoft, Redbubble, TikTok, and Twitter. All signatories commit to enacting safeguards, disrupting harmful advertising, ensuring platform security, empowering users, increasing transparency in political advertising, supporting research, and publicizing their efforts against mis- and disinformation.

Sources: Oversight Board. European Commission. OECD.

Cyberbullying and Online Harassment

Online abuse is an umbrella term used to describe a "diversity of tactics and malicious behaviors ranging from sharing embarrassing or cruel content about a person to impersonation, doxing, stalking and electronic surveillance to the nonconsensual use of photography and violent threats."[55] This term also encompasses cyber bullying and online harassment.[56]

Cyberbullying and online/cyber harassment are not new occurrences and are two growing harmful tactics and behaviors deterring individuals from freely expressing their thoughts and opinions online. Online platforms have provided cyber bullies and harassers with the means to target individuals and online users directly, publicly, and permanently, amplifying the ongoing impact of these negative behaviours. Increasingly, **young people, children, women, and sexual minorities** are facing rising incidents of cyber bullying and harassment online, globally.

Digital anonymity, which is integral to FoE,[57] permits cyber bullies and harassers to hide behind a cloak of anonymity. Further, law enforcement officials struggle to recognize the severity and impact of online abuse.[58] This makes it difficult to hold cyber bullies and harassers to account which creates a culture of impunity and entrenches these practices.

Important Note

"Anonymous communication is seen by many as a cornerstone of promoting freedom of speech, expression and privacy on the internet, but it can also be misused to control and abuse people."

Sources: *Australian eSafety Commissioner.*

Cyberbullying: this involves the "posting or sending of electronic messages, including pictures or videos, aimed at harassing, threatening, or targeting another person."[59] Electronic content aimed at bullying individuals online is spread through various online platforms, including social media platforms, gaming sites, chat rooms, blogs, and

instant messaging, amongst others. Statista reports that in two Latin American countries, Brazil and Argentina, parents were aware that their children were being cyber bullied more by their classmates at 53% and 40% respectively.[60] Cyberbullying has serious implications for young people's and children's mental health impacting them in their capacities as victims, perpetrators, or bystanders. Further, cyberbullying impacts young people's and children's digital rights and internet freedoms, specifically FoE online, digital safety, and privacy.

Online/Cyber Harassment: this is defined as the 'use of information and communications technologies by an individual or a group to repeatedly and intentionally cause harm to another person, by humiliating, annoying, attacking, threatening, alarming, offending and/or verbally abusing individuals.'[61] **Online sexual harassment** is used to refer to a wide range of sexual misconduct on digital platforms… [that disproportionately targets]…those who identify as women and/or LGBTQIA+."[62]

Important Note

Online abuse is being exacerbated by machine learning and artificial intelligence technological breakthroughs.

- **Deepfakes**: these "use AI, specifically deep learning, to create manipulated content (an image, audio, video) that convincingly alters and misrepresents someone as doing or saying something that was not actually done or said."[63] Women are mostly impacted by deepfakes, which continues to be used as a tactic to discredit, silence, and shame women in online public spaces.

Sources: *Canadian Global Affairs Institute*. *Centre for International Governance Innovation*.

Annex 1: Contemporary Issues impacting FoE Online

Contemporary Issue	Impact on FoE
Artificial Intelligence	**AI-powered content filtering systems:** potential for misuse to block legitimate content that can lead to the censoring or silencing of dissenting voices. **AI-powered surveillance systems:** potential for chilling effect on FoE. **Algorithmic bias and manipulation:** Biased which can amplify certain voices or perspectives while suppressing others, leading to a distortion of free expression. **Sources:** Geneva Internet Platform. OHCHR.
Privacy and Data Protection	Collection and processing of vast amounts of personal data by state and non-state actors, and the protection of privacy in a manner that infringes legitimate FoE online. **Sources:** Privacy International. OHCHR. UNESCO.
Digital Divide, Digital Inclusion and Participation	The digital divide is a core barrier to individuals' ability to exercise their FoE online. Access and affordability issues are central to digital divide challenges but fail to address the full range of connectivity challenges faced by groups. In this context, digital inclusion and participation need to be prioritised to create a safe, accessible environment for the meaningful and full exercise of FoE online. **Sources:** Centre for Digital Society. UN Women. IREX. DW.
Digital Literacy and User Awareness	Digital literacy and user awareness empower individuals to navigate online spaces, cognisant of their operating environments (*e.g., monitoring, surveillance*), their own actions (*e.g., responsibilities as users of a shared global space*), and the action/inaction of other stakeholders (*e.g., other users, online communities, state and private actors*). **Sources:** OHCHR. UNESCO.
Global Internet Governance	Global internet governance efforts shape policies, regulations, and infrastructure that can either support or restrict online speech. **Sources:** CoE.WSIS.IGF. UN.

Supplementary Resources

Legal Instruments (International, Regional, National)

African Charter on Human and Peoples' Rights (ACHPR), 1986.

American Convention on Human Rights (1969).

American Declaration of the Rights and Duties of Man (1948).

Association of Southeast Asian Nations (ASEAN) Human Rights Declaration, 2009.

Charter of Fundamental Rights of the European Union, 2009.

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, or ECHR), 1953.

Declaration of Principles on Freedom of Expression in Africa, revised in 2019.

Guidelines on Freedom of Association and Assembly in Africa, 2017.

Inter-American Democratic Charter (2001).

International Covenant on Civil and Political Rights (ICCPR), 1996.

Universal Declaration of Human Rights (UDHR), 1948.

Selected Resources: Online Censorship

Andy Greenberg. The Ingenious Way Iranians Are Using Satellite TV to Beam in Banned Internet.

ARTICLE 19 (2018). Side-stepping rights: Regulating speech by contract.

Committee to Protect Journalists. Critics are not criminals.

CompariTech (2023). Internet Censorship 2023: A Global Map of Internet Restrictions.

Electoral Institute for Sustainable Democracy in Africa (2022). Digital Censorship and Africa's Democratic Future.

Engage Media (2022). Reports: Monitoring the state of internet censorship in South and Southeast Asia.

Ivana Vojinovic (2023). Internet Censorship: Definition, Types, and How It Can Affect You.

Media Defence. Censorship by Private Actors.

Open Technology Fund (2023). The Decentralised Infrastructure of Online Censorship in Asia.

Rochelle Terman. Internet Censorship (Part 2): The Technology of Information Control.

Stanford PACS (2020). Internet Infrastructure and Human Rights: A Reading List.

UNESCO (2023). Safeguarding freedom of expression and access to information: guidelines for a multistakeholder approach in the context of regulating digital platforms.

Selected Resources: Hate Speech and Information Disorders

Colombia University. Case Law on Hate Speech: 1964 – 2023.

The Rabat Plan of Action.

The US Department of Justice. Hate Crimes Case Examples.

Zachary Laub. Hate Speech on Social Media: Global Comparisons.

Selected Resources: Online Abuse

American Library Association (2019). Libraries respond: cyber-bullying and doxing.

BBC. How online gaming has become a social lifeline.

Fereshteh Naseri et. al., A Comparative Study on the Opportunities and Threats of the Internet and Considering the Rights of Kids Online in Australia, Brazil, Iran, and South Africa.

Janis Wolak et al., Does Online Harassment Constitute Bullying? An Exploration of Online Harassment by Known Peers and Online-Only Contacts.

Karen Brown et al., Cyber-Bullying: Developing Policy to Direct Responses that are Equitable and Effective in Addressing this Special Form of Bullying.

Maral Dadvar et al., Cyberbullying detection: a step toward a safer Internet yard.

Pew Research Center. Online harassment occurs most often on social media, but strikes in other places, too.

Robert Tokunaga et al., Cyber-defense: A Taxonomy of Tactics for Managing Cyberstalking.

Sarah Jameson. Cyberharassment: Striking a Balance between Free Speech and Privacy.

Stine Eckert. Fighting for recognition: Online Abuse of Women Bloggers in Germany, Switzerland, the UK and US.

Viktorya Vilk. You're not powerless in the face of online harassment.

Zahra Ashktorab. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers.

Zinar Ghasem et. Al., A hybrid approach to combat email-based cyberstalking.

General Guides

Center for International Media Assistance (CIMA) and the National Endowment for Democracy. 2018. International standards on freedom of expression: A basic guide for legal practitioners in Latin America and the Caribbean.

Colombia University – Global Freedom of Expression. Case Law.

The Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda (2017 Joint Declaration).

OHCHR. Universal Human Rights Index.

OHCHR. Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression A/HRC/29/32.

OHCHR. Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression, on freedom of expression A/HRC/32/38.

OHCHR. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression A/HRC/35/22.

OHCHR. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression A/HRC/38/35.

Global Events

Access Now. RightsCon

ARTICLE 19. Internet Freedom Festival.

Engage Media. Digital Rights Asia-Pacific 2023

UN. Internet Governance Forum.

Collaboration on International ICT Policy for East and Southern Africa (CIPESA). Forum on Internet Freedom in Africa (FIFAfrica) 2023.

Freedom Online Coalition. Freedom Online Conference.

References

[1] IGI Global. What is Internet Censorship?
[2] Council of Europe. Information Disorder.
[3] US Department of State. Internet Freedom Fact Sheet.
[4] Universal Declaration of Human Rights (UDHR), 1948. International Covenant on Civil and Political Rights (ICCPR), 1966.s
[5] OHCHR, Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework (The Ruggie Principles), A/HRC/17/31, 21 March 2011, Annex. The UN HRC endorsed the Guiding Principles in HRC Resolution 17/4, A/HRC/RES/17/14, 16 June 2011.
[6] UN Human Rights Council. Resolution 20/8 on the promotion, protection and enjoyment of human rights on the Internet, A/HRC/RES/20/8.
[7] Article 19 (2) of the ICCPR.
[8] UN Human Rights Committee. General Comment No. 34, note 4, para 43.
[9] European Court of Human Rights. *Thorgeirson v. Iceland*, Application No. 13778/88, para. 63.
[10] UN Human Rights Committee. General Comment No. 34, note 4, para 21.
[11] Article 9 of the African Charter on Human and Peoples' Rights.
[12] The Declaration of Principles of Freedom of Expression and Access to Information in Africa. This was adopted by the ACHPR African Commission on Human and Peoples' Rights (the African Commission) in 2019 at its 65th Ordinary Session.
[13] Principle 3 of the African Declaration on Internet Rights and Freedoms.
[14] ARTICLE 19. Central Asia: Freedom of Expression Online.
[15] OHCHR. Central Asia Regional Office.
[16] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, or ECHR), 1953.
[17] UNESCO (2015) Freedom of Expression and Public Order: Training Manual.
[18] This list is not exhaustive and also include the denial of historical events and laws that protect the State, state figures or state symbols from insult or criticism, provided these are not captured by the Article 19 (2) and Article 20 (2) provisions.
[19] *Handyside v. the United Kingdom*, 1976. Application no. 5493/72
[20] ARTICLE 19. Hate Speech Explained: A Summary.
[21] OHCHR. A/HRC/47/25: Disinformation and freedom of opinion and expression - Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.
[22] Cambridge Dictionary. Satire.
[23] Cambridge Dictionary. Parody.

[24] ARTICLE 19. Hate Speech Explained: A Summary.
[25] UN Special Rapporteur on the right to freedom of religion or belief. Elimination of all forms of religious intolerance. See also: UNHRC. General Comment No. 34, p 12.
[26] HumanistsUK. Abolition of English and Welsh Blasphemy Laws. The Guardian. Denmark Scraps 334-year old Blasphemy Law. End Blasphemy Laws. Canada Repeals "Blasphemy" Laws.
[27] End Blasphemy Law. Ethiopia.
[28] OHCHR, Vitit Muntarbhorn. Study on the prohibition of incitement to national, racial or religious hatred: Lessons from the Asia Pacific Region.
[29] US Department of State. 2022 Report on International Religious Freedom for Saudi Arabia.
[30] Examples include intergovernmental coalitions, such as the Freedom Online Coalition, the multistakeholder coalitions based at the UN Internet Governance Forum, the Internet Rights and Principles Coalition. Regional coalitions include the African Declaration on Internet Rights and Freedoms and the African Internet Rights Alliance (AIRA) (African region); the SEE Digital Rights Network covering the Balkans and Central Asian regions; the Alliance for Encryption in Latin America and the Caribbean (AC-LAC) (Latin America and the Caribbean region), and the Asia Internet Coalition and the Southeast Asia Freedom of Expression Network (SSE Asia region).
[31] A few instances of collaboration are evidenced in the following documents: OHCHR, Special Rapporteur on Freedom of Expression and Opinion. Joint Declarations.
[32]
[33] 'What is Internet Censorship?' (IGI Global) https://www.igi-global.com/dictionary/internet-censorship-china/42010.
[34] Stephen A. Meserve and Daniel Pemstein. Google Politics: The Political Determinants of Internet Censorship in Democracies.
[35] Catherine Andrzejewski, Ana Horigoshi, Abigail I. Maher, and Jonathan A. Solis. Innovators and Emulators: China and Russia's Compounding Influence on Digital Censorship.
[36] Margaret E. Roberts. Resilience to Online Censorship. Also: Freedom House. Freedom on the Net 2022: Countering an Authoritarian Overhaul of the Internet.
[37] Internet shutdowns, surveillance and privacy topics are explored separately under Topic 1 and Topic 6 respectively.
[38] Electronic Frontier Foundation. Surveillance Self-Defense. Freedom House. Leaping Over the Firewall: A Review of Censorship Circumvention Tools. Electronic Frontier Foundation. Yi Mou, Y, Kevin Wu, & David Atkin. Understanding the use of circumvention tools to bypass online censorship. The Citizen Lab. Everyone's Guide to Bypassing Internet Censorship. OpenNet Initiative. About Filtering.

[39] See: The Ranking Digital Rights. The 2022 Telco Giants Scorecard. Also: Prem M. Trivedi. Content Governance in the Shadows: How Telcos & Other Internet Infrastructure Companies "Moderate" Online Content.
[40] This joint call was framed in the context of content moderation but has general applicability.
[41] The Association for Progressive Communications. Freedom of expression and the private sector in the digital age - Submission to the United Nations Special Rapporteur on the Right to Freedom of Opinion and Expression by the Association for Progressive Communications (APC). See also: Corinne Cath, Niels ten Oever and Daniel O'Maley. Media Development in the Digital Age: Five Ways to Engage in Internet Governance.
[42] Prior restraint is "a form of censorship that allows the government to review the content of printed materials and prevent their publication." See: Daniel Baracskay. Prior Restraint.
[43] Tech Law. Kenyan court paves way for lawsuit alleging Facebook played role in fuelling Ethiopia's Tigray conflict. Medium. Influential Ethiopian social media accounts stoke violence along ethnic lines. OHCHR. Myanmar: Social media companies must stand up to junta's online terror campaign, say UN experts.
[44] OHCHR. Moderating Content Online: Fighting Harm or Silencing Dissent.
[45] Trust and Safety Professional Association. What is Content Moderation?
[46] Devi Soni. Machine Learning for Content Moderation -Challenges.
[47] Kai Riemer and Sandra Peter. Wrong, Elon Musk: the big problem with free speech on platforms isn't censorship. It's the algorithms.
[48] European Parliament. The impact of algorithms for online content filtering or moderation.
[49] Christopher S. Leberknight and Mung Chiang. A Taxonomy of Internet Censorship and AntiCensorship.
[50] OHCHR. Report: Online hate increasing against minorities, says expert.
[51] Matthew L Williams and others. Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime.
[52] *Ibid, n. 50.*
[53] Zachary Laub. Hate Speech on Social Media: Global Comparisons.
[54] Claire Wardle and Hossein Derakhshan, '*Information Disorder: Toward an interdisciplinary framework for research and policy making*' Council of Europe 2017.
[55] Women's Media Centre. Online Abuse 101.
[56] PEN America. Defining "Online Abuse": A Glossary of Terms.
[57] OHCHR. Human rights, encryption and anonymity in a digital age.
[58] Media Defence. Online Harassment.
[59] UN. Bullying and Cyberbullying.

[60] Statista. Knowledge of the type of relationship children victims of cyber bullying had with their harassers according to parents in Brazil as of April 2018. Statista. Knowledge of the type of relationship children victim of cyber bullying had with their harassers according to parents in Argentina as of April 2018.
[61] UNODC. Cyberstalking and Cyberharassment; Durham University. What is Online Harassment?
[62] *Ibid, n. 56.*
[63] Merriam-Webster. Deepfakes. CGIA. The Use and Abuses of Deepfake Technology.