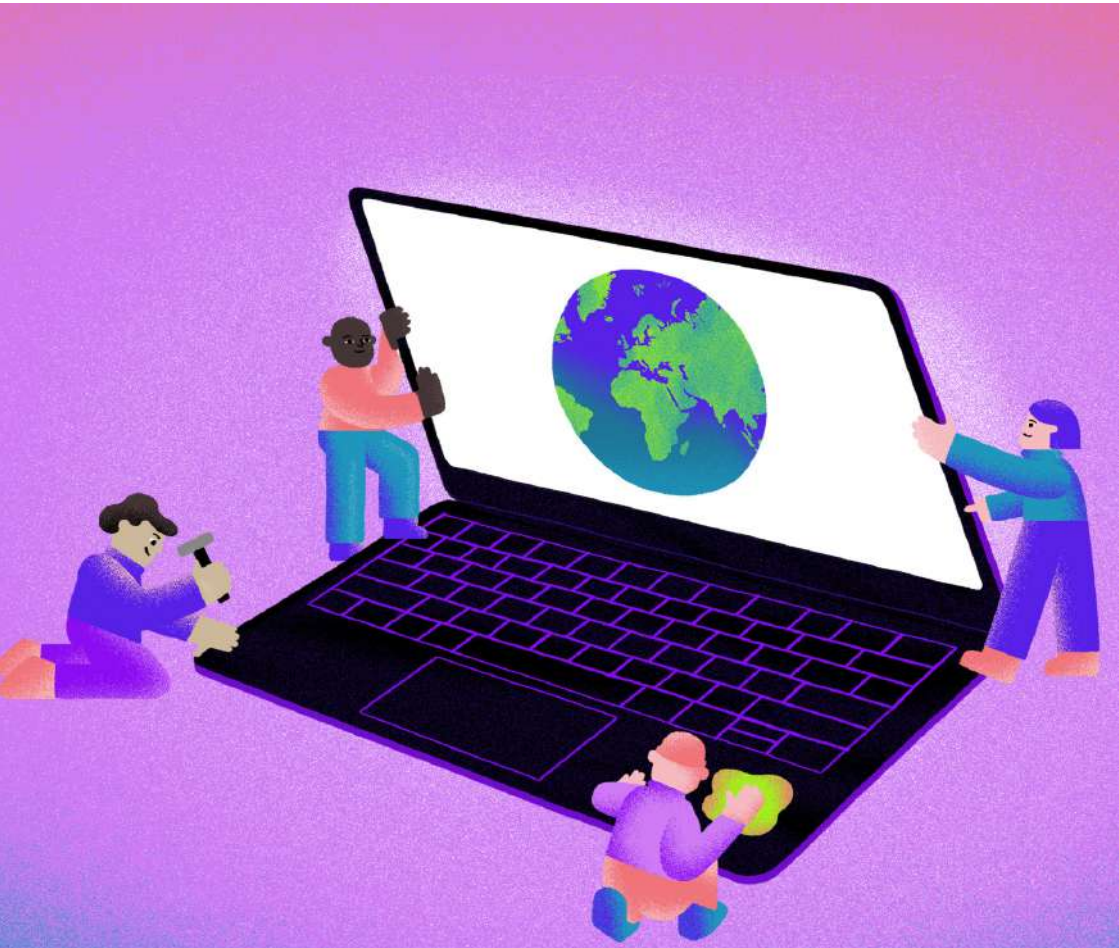
CDJMC, TU

A Toolkit for Civil Society on Digital Platforms Governance



Published under the initiative

"Safeguarding Freedom of Expression and Access to Information through the Implementation of the UNESCO Guidelines for the Governance of Digital Platforms,"

Jointly undertaken by
Tribhuvan University, Nepal; Jahangirnagar University, Bangladesh;
and the University of Vocational Technology, Sri Lanka,
with the support of UNESCO.

April 2026

A Toolkit for Civil Society on Digital Platforms Governance

ISBN: 978-9905-0-1074-1 (Print version)

ISBN: 978-9905-0-1108-3 (e-Book)

Edition: 1st | Print: 1000 Copies

This Toolkit is available at: cdjmc.tu.edu.np

© UNESCO/CDJMC, TU

Cover, Typesetting and Graphic Design: UNESCO Template

Published by:

Central Department of Journalism and Mass Communication
(CDJMC), Tribhuvan University, with the support of UNESCO.

Layout & Typesetting: Jit B. Pun
Printed by: Thememaker Pvt. Ltd, Adwait Marg, Kathmandu

A Toolkit for Civil Society on Digital Platforms Governance

unesco
United Nations
Educational, Scientific
and Cultural Organization

UNIVOTEC

TABLE OF CONTENTS

Contents

5 **Foreword**

7 **Preface**

11 **Executive Summary**

13 **Introduction**

26 Supporting Freedom of Expression and Access to Information

33 Addressing Harmful Content and Identifying Roots of Abusive Behaviours

47 Fact-Checking and Watchdogging Digital Platforms

54 Promoting Transparency and Accountability of Digital Platforms

61 User Grievance Mechanisms and Independent Oversight

68 Conclusion: Strengthening Civil Society Engagement in Digital Platform Governance

71 Way Forward: Practical Guide to Integrating the Toolkit for Civil Society

76 Annex 1

85 Annex 2

87 Annex 3

89 References

Foreword

Across South Asia, digital platforms have opened extraordinary spaces for public expression and have become sites of contest over who may speak, who is heard, and who is silenced. Civil society organizations, journalists, fact-checkers, researchers and community groups are usually the first to see how these dynamics unfold: how disinformation spreads in languages that receive little moderation attention, how abuse falls disproportionately on women and minorities, how users are left without meaningful redress. They are also best positioned to document these harms, to demand accountability, and to help people navigate the information environment.

UNESCO's Guidelines for the Governance of Digital Platforms, published in November 2023 following a consultation that gathered more than 10,000 comments from 134 countries, recognize this explicitly. The governance of digital platforms cannot rest with States or platforms alone. The Guidelines set out duties, responsibilities and roles for civil society, media, academia and the technical community alongside those of governments and companies, because an "Internet of Trust" is a shared undertaking that requires a whole-of-society effort in which independent voices are not merely consulted, but structurally engaged.

This Toolkit for Civil Society is designed to make that engagement concrete. It moves from principle to practice across seven units, where each unit offering action points, implementation checklists and monitoring indicators — tools that enable organizations of offering different sizes and mandates to act on evidence rather than impression.

The Toolkit is built from the realities of Bangladesh, Nepal and Sri Lanka, drawing on interviews and workshops held in Colombo, Dhaka and Kathmandu between December 2025 and March 2026 with more than 300 contributors and significant participation from women, alongside a systematic review involving relevant laws and policies, and subsequently validated at a regional workshop in Kathmandu in March 2026. The Toolkit also advances to UNESCO's regional priorities for South Asia, by focusing on rights-based digital platform governance, media and information literacy, transparency, accountability and user redress. It supports UNESCO's broader commitment to digital transformation, gender equality, youth engagement and inclusive participation across the region.

UNESCO expresses its appreciation to the European Union, the participating universities and all partners and stakeholders who contributed to this important regional effort.

This Toolkit belongs to those who will use it; and UNESCO remains committed to working alongside civil society in South Asia to advance freedom of expression, access to information, and building a digital environment that is open, safe, inclusive and grounded in human rights.

Jaco du Toit
UNESCO Representative to Nepal and Head of Office

Preface

Indeed, the Toolkit for Civil Society, developed to support the implementation of the UNESCO Guidelines for the Governance of Digital Platforms in the South Asian context, is now in your hands. On behalf of the Departments of Media Studies at Jahangirnagar University, Bangladesh; the University of Vocational Technology, Sri Lanka; and Tribhuvan University, Kathmandu, Nepal, I am pleased to present this Toolkit as the outcome of extensive observations of the digital platform landscape, consultations with users and policymakers, and a review of relevant legal frameworks across these three countries.

The project, Safeguarding Freedom of Expression and Access to Information through the Implementation of the UNESCO Guidelines for the Governance of Digital Platforms, was initiated in August 2025. At the outset, only a preliminary framework was available, and we planned to employ multiple approaches to map the existing landscape of digital platform governance across Bangladesh, Nepal, and Sri Lanka. We closely examined similarities in their socio-political contexts, including growing frustration among youth over governments' inability to adequately protect citizens' right to freedom of expression. We also engaged with diverse stakeholders—including government officials, politicians, media professionals, academics, and members of civil society—to address an important question: How can the rights and freedoms of citizens be better protected within the contemporary digital media environment?

The process began with individual consultations and subsequently expanded through workshops conducted in Colombo, Dhaka, and Kathmandu, with significant participation from women. In parallel, country briefs were prepared to assess the legal provisions governing digital platforms in Bangladesh, Nepal, and Sri Lanka. The Toolkit evolved through an empirical process involving both individual and group consultations conducted through interviews and workshops between December 2025 and March 2026. More than 300 individuals contributed their expertise and perspectives throughout this process.

Simultaneously, a systematic review of relevant laws and policies was undertaken and played a crucial role in shaping the Toolkit. Insights generated through these engagements were further refined through continued discussions with representatives from law enforcement agencies, ministries of law and communication, members of parliament, political actors, media organizations, and civil society representatives from diverse sectors. The drafting process continued until the final stage to incorporate perspectives from government agencies, lawmakers and political actors, digital platforms and telecommunications authorities, intergovernmental institutions such

as Press Councils and Information Commissions, journalists' unions, academia, media organizations, and the technical community. We therefore believe that this Toolkit represents the outcome of a genuinely inclusive and multi-stakeholder consultation process across Bangladesh, Nepal, and Sri Lanka.

Before its finalization, the Toolkit was validated and refined through structured feedback received during a regional workshop held in Kathmandu in March 2026. The workshop brought together a diverse range of stakeholders from Bangladesh, Nepal, and Sri Lanka, as well as observers from the Maldives. Participants included policymakers, such as Members of Parliament, alongside officials from UNESCO and I4T at both headquarters and country levels.

The Toolkit embodies the spirit of the Pretoria Proposition for Action (2026), which recognizes that effective governance cannot rest solely on digital platforms themselves or on regulators acting independently. The Proposition further emphasizes that civil society organizations, researchers, educators, and affected communities play a critical role in identifying systemic risks, supporting redress mechanisms, and strengthening media and information literacy.

An important aspect worth highlighting is that the Toolkit also outlines several future steps that UNESCO may consider operationalizing in the years ahead. It further includes provisions for institutional adoption and curriculum integration within universities. Designed specifically for civil society organizations and actors advocating for human rights in the digital sphere, the Toolkit aims to create meaningful opportunities for sharing experiences, good practices, and challenges while contributing to a coordinated and multi-stakeholder regional approach.

The practical checklist and monitoring guide developed within the Toolkit are designed to strengthen users' resilience and empower communities to respond effectively to online harm. They can serve as important resources for civil society organizations seeking to promote accountability and transparency in digital platforms in a systematic manner. The subsequent units and case studies included in the annexes provide practical and engaging examples that help civil society actors better understand the issues and indicators they are expected to examine.

Careful attention has also been given to the structure and presentation of the Toolkit to ensure that it remains both comprehensive and accessible. It broadly follows American Psychological Association (APA) style conventions, and necessary explanations and contextual clarifications have been integrated into the text where appropriate to enhance readability.

At this moment of successfully completing this initiative and presenting the Toolkit for Civil Society, I would like to express my sincere gratitude to UNESCO officials, from the Kathmandu Office to Headquarters. I would particularly like to acknowledge the support and encouragement provided by Mr. Jaco du Toit, UNESCO Representative to Nepal and Head of Office, and his colleagues, including Ms. Nirjana Sharma, Lead for the Communication and Information Sector/Programme Officer; Ms. Aruna Joshi, Digital Media Associate; and Ms. Yujie Fan, Trainee. Their unwavering support throughout the preparation process significantly contributed to the completion of this Toolkit.

I would also like to express my sincere appreciation to Ms. Ana Cristina Ruelas, Senior Programme Specialist, Section for Freedom of Expression and Safety of Journalists, Communication and Information Sector, UNESCO; and to Mr. Christophe Gauthier and Mr. Chandra Bhanu Patnaik, Co-Chairs of the I4T Knowledge Network, for their valuable technical expertise, without which this Toolkit would not have taken its present shape.

The preparation and publication of this toolkit would not have been possible without the invaluable contributions, hard work, and dedication of H.A. Gayan Madushanka, Head of the Department of Film & Television Production Technology at the University of Vocational Technology, Sri Lanka, and Rakib Ahmed, Associate Professor in the Department of Journalism & Media Studies at Jahangirnagar University, Bangladesh. Both colleagues have been closely involved in this initiative from its inception. They played a key role in organizing the respective national consultation workshops in Sri Lanka and Bangladesh, actively participated in the regional workshop in Kathmandu, and contributed significantly to the development and refinement of the toolkit's content. Their expertise, thoughtful insights, and sustained commitment substantially enriched the intellectual and practical dimensions of this publication. This Toolkit is the outcome of a rigorous collaborative effort involving experts and stakeholders from various countries and institutions, including UNESCO and the I4T initiative. I gratefully acknowledge Dr. Ghama Raj Luitel's contribution throughout the process, from research to the compilation of materials. I sincerely appreciate their valuable support and contributions throughout the process.

Last but not least, I would like to thank my team, including faculty members at the Central Department of Journalism and Mass Communication: Dr. Shree Ram Paudel, Dr. Rishikesh Dahal, Dr. Deb Raj Aryal, and Dr. Shashi Marasini. I also extend my gratitude to experts beyond the Department, including Mr. Ujjwal Acharya, Mr. Santosh Sigdel, and Mr. Tara Nath Dahal, Hiroyuki Tiara as well as the administrative staff of the Department, Ms. Kalpana Bajracharya and Mr. Eka Raj Kadel. I would like to express my sincere thank to Ms. Iksha Limbu, an MPhil-PhD scholar at CDJMC,

for her careful and meticulous final proofreading. I firmly believe that without the dedication and hard work of all these individuals, the publication of this Toolkit would not have been possible.

Kundan Aryal, PhD
Professor and Head of the Department
Central Department of Journalism and Mass Communication
Tribhuvan University, Nepal

Executive Summary

This Toolkit aims to provide practical guidance for operationalizing the UNESCO Guidelines for the Governance of Digital Platforms in Bangladesh, Nepal, and Sri Lanka. While developed in the context of these three countries, it may also be broadly applicable across South Asia and other regions with similar socio-political and economic contexts. Since it identifies eight common concerns regarding digital platforms across these countries, many of these issues may also be relevant to other parts of South Asia and the wider world. This initiative has been undertaken as a collaborative effort between UNESCO and partner universities under the project titled *"Safeguarding Freedom of Expression and Access to Information through the Implementation of the UNESCO Guidelines for the Governance of Digital Platforms."*

The introductory section of the Toolkit explains its significance, outlines its objectives, and provides guidance on how to use it effectively. The introductory section is followed by seven thematic units designed to facilitate practical implementation. The Toolkit is comprehensively grounded in five key principles: platforms conduct human rights due diligence; platforms adhere to international human rights standards, including in platform design, content moderation, and content curation; platforms maintain transparency; platforms provide users with access to relevant information and tools; and platforms remain accountable to relevant stakeholders.

The Toolkit is primarily designed for civil society organizations, including media institutions, fact-checking organizations, academia and academic research institutions, digital rights organizations, and community-based organizations. It defines both civil society and digital platforms within the broader context of governance and participation. Accordingly, its major objective is to foster an environment in which freedom of expression and access to information are placed at the core of digital platform governance processes, while defining the duties, responsibilities, and roles of civil society, media, academia, the technical community, and other stakeholders, as outlined in the UNESCO Guidelines.

As the Toolkit was developed through an empirical process involving individual and group consultations in the form of interviews and workshops conducted across the three countries between December 2025 and March 2026, it reflects the realities of digital platform governance in Bangladesh, Nepal, and Sri Lanka. Altogether, more than 300 individuals contributed their opinions and expertise during the consultation phase. Simultaneously, a systematic review of relevant laws and policies was undertaken, which played an instrumental role in shaping the Toolkit. Before being finalized, the Toolkit was validated and refined during a regional workshop organized in Kathmandu

in March 2026, attended by a diverse range of stakeholders from Bangladesh, Nepal, and Sri Lanka, including policymakers such as Members of Parliament, as well as officials from UNESCO and I4T at both headquarters and country levels.

Following the introductory section, the Toolkit is structured around seven units that progressively move from practical approaches to safeguarding freedom of expression and access to information to broader strategies for strengthening digital platform governance. It begins with hands-on methods to support freedom of expression and access to information (Unit I), followed by approaches for addressing harmful content and identifying the roots of abusive behaviour (Unit II). The Toolkit then presents methods for fact-checking and monitoring digital platforms (Unit III), along with practical guidance for promoting transparency and accountability within digital platforms (Unit IV).

Subsequent units focus on actionable measures for civil society to enhance the effectiveness of user grievance mechanisms and independent oversight systems (Unit V), and on strengthening civil society engagement in digital platform governance (Unit VI). Finally, the Toolkit concludes with a forward-looking perspective (Unit VII), which serves as a practical guide for integrating its principles and approaches into university curricula and other institutional settings.

The Toolkit is particularly well suited for educational purposes. Its clearly structured units, supported by practical scenarios, checklists, tools, and monitoring indicators for harmful content, provide a useful framework for learning and application. Furthermore, its strong emphasis on human rights due diligence, multi-stakeholder approaches, transparency, and systemic oversight, rather than heavy-handed content policing, makes it highly relevant for academic and professional contexts. It also aligns closely with UNESCO's broader initiatives on Media and Information Literacy (MIL), AI competency frameworks for students and teachers, and the ethical integration of artificial intelligence in higher education. In addition to the thematic units, the annexes provide key terminologies as well as relevant case studies from different countries, highlighting good practices and practical examples.

Overall, the Toolkit envisions a proactive and constructive role for civil society in strengthening trust in the Internet and fostering a more inclusive and rights-based digital environment. The Toolkit comprehensively and coherently reflects UNESCO's broader objectives of ensuring that freedom of expression, access to information, and diverse cultural content are fully protected for all. At the same time, it recognizes the need for collaborative efforts among various stakeholders, including government agencies, to address challenges such as disinformation, misinformation, and online hate speech.

Introduction

Digital platforms have transformed global communication, information dissemination and public discourse. Therefore, they play a major role in shaping the information ecosystem, public discourse and democratic processes. These platforms now also play a central role in shaping political debates, cultural interactions and economic activities. While these platforms have expanded opportunities for expression and participation, they have also generated serious concerns regarding their effects on democratic discourse. Such concerns comprised of content related harms such as misinformation and hate speech as well as dissemination mechanisms such as algorithmic manipulation.

In response to these challenges, the United Nations Educational, Scientific and Cultural Organization (UNESCO) introduced the Guidelines for the Governance of Digital Platforms (UNESCO, 2023). This toolkit is prepared to contextualize the guidelines, which were developed through extensive global consultations involving governments, technology companies, civil society organizations, academic institutions and independent experts. The purpose of the toolkit is to support the implementation of the guidelines in Bangladesh, Nepal and Sri Lanka by advancing their central objective: to establish a human rights friendly framework for governing digital platforms that safeguards freedom of expression while effectively addressing online harms.

The toolkit for civil society is the result of consultations in these three nations, enriched by national workshops and by comments and feedback collected from over 300 stakeholders in the context of the subregion. This toolkit reflects inputs from different stakeholders from Bangladesh, Nepal and Sri Lanka.

Why this toolkit?

This toolkit aims to support civil society organizations with the knowledge, frameworks and practical tools essential to engage and participate in the governance of digital platforms in line with the UNESCO Guidelines for the Governance of Digital Platforms. It further aims to support civil society to monitor platform practices, advocate for human rights–based digital policies, address harmful online content and promote transparency, accountability and user empowerment.

By strengthening the capacity of civil society organizations, including non-governmental organizations, rights advocates, media and academia, the toolkit supports multistakeholder collaboration and encourages the development of a trustworthy and rights-respecting digital information environment in Bangladesh, Nepal and Sri Lanka.

Overall, this toolkit aims to provide practical guidelines for civil society to safeguard freedom of expression, protect users from online harms, participate meaningfully as a stakeholder in digital platforms governance and contribute to accountable digital platform governance in the region.

Digital platforms have become central actors in modern information ecosystems. Their algorithms determine which information is amplified, which voices are prioritized or deprioritized, and how users interact with content online. As a result, platforms hold significant influence over the formation of public opinion and democratic discourse. However, the rapid growth of these platforms has also revealed structural challenges. Misinformation has affected elections and public discourses. Algorithmic recommendation systems may amplify sensational or harmful content in order to maximize user engagement. Furthermore, the concentration of power in a few major technology companies has raised concerns about transparency, accountability and the protection of fundamental rights.

Governments around the world have attempted to address these issues through various regulatory initiatives. Yet some regulatory efforts risk undermining freedom of expression or enabling censorship. Recognizing this paradoxical situation, UNESCO developed the Guidelines to provide a human-rights-based framework for digital platform governance. The framework aims to help states regulate digital platforms responsibly while ensuring that fundamental freedoms remain protected. Another key rationale behind the development of guidelines is to address the challenges created by the global nature of digital platforms. Since online services operate across national borders, isolated national regulations are often insufficient. The UNESCO guidelines therefore encourage regional/international cooperation and common standards that can guide policy development worldwide.

Specific objectives of the toolkit

Civil society, media, academia, the technical community and other stakeholders play a critical role in promoting human rights and accountability in the digital platform governance system (Article 19, 2020). By monitoring and reporting online harms, disinformation, and content moderation practices, they help ensure that platforms operate transparently and responsibly. CSOs can also support capacity building for regulators, law enforcement agencies, and legal professionals on harmful content, digital safety, and rights-based platform governance. As the role of the media is crucial in spreading media and information literacy, academia present findings on anomalies or recommend areas for improvement.

This toolkit is designed for:

- **Civil society organizations**
- **Media**
- **Fact-checking organizations**
- **Academia and academic research institutions**
- **Digital rights organizations**
- **Community organization**

The civil society serves as independent watchdogs, providing evidence-based insights, raising public awareness and advocating for the rights of marginalized communities, religious minorities, journalists, human rights defenders and other at-risk groups.

Beyond advocacy and oversight, these stakeholders contribute technical and research expertise to strengthen digital governance. They conduct studies on platform impacts, algorithms, and user behaviour to inform policies and platform practices, participate in advisory committees, task forces, and working groups, and provide guidance on accessibility, inclusivity, and ethical technological solutions. Their active engagement ensures that decision-making processes are balanced, transparent, and aligned with international human rights standards.

Thus, the objectives of this toolkit will be best achieved when civil society, including the media, academia the technical community, and other stakeholders in the subregion, plays a more instrumental role in holding states and digital platforms accountable for digital rights and the full spectrum of interconnected human rights.

How to use this toolkit?

This toolkit can be used to:

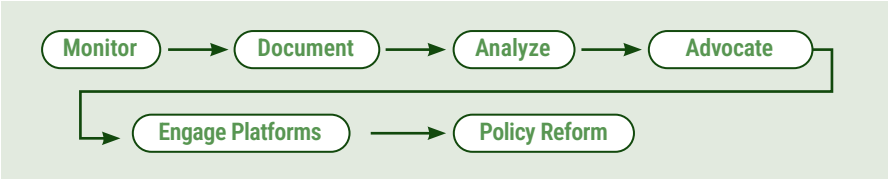
- **monitor digital platform practices**
- **conduct digital platform research**
- **document online harms**
- **engage in policy advocacy**
- **support victims of online abuse**
- **participate in multistakeholder governance processes**
- **promote media and information literacy**

This toolkit also serves as a practical manual for civil society organizations to navigate digital platform governance as it includes unit-based checklists to design activities, conduct research or advocacy initiatives along with templates and examples. Some suggestive practical uses of this toolkit are as follows:

- **Fact-checking and watchdogging resource:** Draw from Unit IV and the guide such as 'Steps in Fact-Checking' to strengthen information integrity and hold platforms accountable for misinformation.
- **Monitoring framework application:** Use the 'Monitoring Framework for Harmful Content' (Unit II) to identify threat actors, affected groups and the impact of algorithmic amplification in local context.
- **Support and legal guidance:** Apply strategies from Units II and III to develop helplines, referral networks and psychosocial support for victims of online abuse.
- **Policy advocacy tool:** Leverage regional insights and UNESCO principles to engage governments and regulators, ensuring laws align with international human rights standards.
- **Media and Information Literacy capacity building:** Use the 'Steps to Build Digital Awareness' (Unit V) to design educational programs for youth and marginalized communities.
- **Serve as user Grievance Mechanisms and Independent Oversight (Unit VI).**
- **Serve the tool for strengthening Civil Society Engagement in Digital Platform Governance (Unit VII).**
- **Integration of the toolkit with the universities' curricula:** Draw from unit VII the toolkit would be instrumental to provide a practical guide to integrate the quest of digital platform governance.

This cycle below shows how civil society organizations can systematically promote accountability in digital platform governance. The process begins with monitoring, where organizations observe platform practices, policies and emerging online harms. Findings are then documented through evidence collection and research reports. The next step is analysis, where patterns, impacts and systemic risks are examined to understand the broader implications. Based on this evidence, civil society can advocate for change through public awareness, policy recommendations and campaigns. This is followed by engagement with platforms, where findings are communicated to companies to encourage improvements in policies, moderation practices and

transparency. Finally, these efforts contribute to policy reform, influencing laws, regulations, and governance frameworks.



UNESCO Guidelines for Digital Platform Governance: An Overview

The UNESCO Guidelines for the Governance of Digital Platforms provide a global framework to help governments, digital platforms, regulators, civil society and other stakeholders develop governance systems that protect freedom of expression, access to information and human rights online. The guidelines promote a multistakeholder, transparent and rights-based approach to digital platform governance.

Key principles of the guidelines include:

- **Human rights–based governance:** Platform policies, curation, and moderation practices should respect international human rights standards, particularly freedom of expression and non-discrimination.
- **Transparency:** Platforms should disclose their content moderation policies, curation, advertising practices and decision-making processes in clear and accessible ways.
- **Accountability:** Platforms should establish effective complaint and redress mechanisms and be accountable to users, regulators and the public.
- **User empowerment:** Users should have access to information about platform rules and be able to understand, question and appeal moderation decisions.
- **Multistakeholder participation:** Policymaking and governance processes should involve governments, platforms, civil society, academia, media and technical experts.

Policy directions encouraged by the guidelines:

- **Promote co-regulatory and self-regulatory mechanisms alongside legislation.**
- **Address online harm such as misinformation, disinformation and hate speech while safeguarding freedom of expression and the right to access information.**
- **Strengthen transparency and oversight mechanisms for platform governance.**
- **Support media and information literacy to help users critically engage with online content.**

Overall, the UNESCO guidelines aim to build an "Internet of Trust", where digital platforms operate transparently, respect human rights, and contribute to a healthy and democratic information ecosystem.

Civil society for Internet of Trust

The UNESCO Guidelines envisages the role of civil society in operationalizing governance processes and policy implementation. It recommends developing digital platform governance policies through open and participatory processes. The process of policymaking should involve consultation with diverse stakeholders, including civil society, and should be based on empirical evidence and research. Civil society is expected to ensure that governance frameworks promote cultural diversity and ensure that digital platforms support a pluralistic information environment where diverse voices and perspectives can be represented.

Amid the reality of governments misusing digital platforms, governance practices such as content takedowns, internet shutdowns and speech-suppressing laws remain prevalent. In this context, civil society requires guidance to effectively monitor, challenge, and intervene against such practices. At the same time, digital platforms often apply stricter moderation standards to English-language content than to content in local languages. This asymmetry has been well documented in South Asia and should be explicitly highlighted. Civil society therefore needs clear monitoring frameworks and advocacy tools to track and respond to these disparities.

The UNESCO Guidelines emphasize that digital platform governance requires collaboration among multiple stakeholders such as the governments, technology companies, civil society organizations, academic institutions and international organizations. In a bid to support the implementation of the Guidelines, this toolkit is focused on the role and responsibilities of civil society to maintain and enhance a healthy digital ecosystem. Civil society organizations and researchers contribute by monitoring

platform practices, conducting independent studies and advocating for user rights. They can play an instrumental role to ensure technology companies, as operators of digital platforms, implement policies aligned with human rights standards and address online harms responsibly.

Within the digital ecosystem, civil society, comprising non-governmental, non-profit, voluntary and independent organizations operating between the state, the private sector and citizens, play a vital role in representing public interests, promoting democratic participation, and advocating for social justice, accountability and human rights. Media and journalism organizations, academic and research institutions, digital rights and human rights organizations, civil liberties and legal advocacy groups, community-based and grassroots organizations, consumer protection and user rights groups, digital literacy and education organizations, youth and student groups, technology policy and internet governance organizations, and organizations representing marginalized or vulnerable communities all have a stake in digital governance. Through research, advocacy, monitoring, public engagement, and policy participation, these actors help ensure that digital governance frameworks and platform practices uphold human rights, transparency, privacy, freedom of expression and inclusive participation, while amplifying the perspectives of diverse and marginalized communities.

Civil society organizations should develop comprehensive policies and internal guidelines for using digital platforms in ways that protect both users and non-users and help mitigate potential harms. The UN document with safety steps to protect children and youth from harmful online content could be an effective tool in this regard. (https://www.un.org/en/global-issues/child-and-youth-safety-online)

Nature of Civil Society

- **Independent and non-governmental:** Civil society organizations operate independently from governments and private corporations while engaging with both to promote public interest.
- **Public interest–oriented:** Their activities focus on protecting human rights, democratic values, social justice, and public accountability.
- **Representative and inclusive:** Civil society often represents the perspectives of diverse communities, including marginalized and vulnerable groups whose voices may be underrepresented in policymaking. Vulnerable groups include women, children, ethnic and linguistic minorities, refugees, persons with disabilities, and neurodivergent individuals.

- **Advocacy-driven:** Many civil society actors work to influence laws, policies, and institutional practices through evidence-based advocacy.
- **Collaborative and network-based:** Civil society frequently works through coalitions, partnerships, and regional or international networks to strengthen impact.
- **Watchdog and accountability role:** They monitor the actions of governments, corporations, and institutions to ensure transparency, ethical conduct, and respect for rights.

What are digital platforms?

In the context of digital platform governance, digital platforms are online services that facilitate interactions among users, businesses and content through digital infrastructure and algorithms (Wardle & Derakhshan, 2017). They are the online systems that organize and distribute information and interactions on the internet, making them central actors in digital communication and public debate.

The UNESCO Guidelines for the Governance of Digital Platforms refers to large online intermediaries that organize and moderate information flows and social interactions online, influencing how information is produced, shared and consumed.

Common characteristics of digital platforms

- **Intermediation:** They act as intermediaries connecting users, creators, advertisers and other actors.
- **User-generated content:** Users can create, upload and share content such as posts, videos, comments, or reviews.
- **Algorithmic curation:** Platforms use algorithms to recommend, rank or prioritize content.
- **Large-scale reach:** They enable a broader of communication and information sharing.
- **Data-driven operations:** Platforms collect and process user data to personalize content and services.

Examples of digital platforms:

- **Social media platforms:** Facebook, X (Twitter), TikTok, Instagram
- **Video-sharing platforms:** YouTube
- **Messaging platforms:** WhatsApp, Telegram
- **Search engines:** Google Search
- **App stores:** Apple App Store, Google Play Store, Samsung Galaxy Store
- **Content sharing and discussion platforms:** Reddit, Discord, forums, blogs
- **Online marketplaces and service platforms:** Pathao, Foodmandu
- **Artificial Intelligence:** ChatGPT, Claude, Google AI Studio, Deepseek, Gemini etc.

Regional Context of Digital Platform Governance: Bangladesh, Nepal, and Sri Lanka

Recent developments in digital platform governance in Bangladesh, Nepal, and Sri Lanka exhibit commonalities. Such commonalities are manifested in regulatory frameworks, institutional arrangements and policy debates related to platform accountability, content moderation and user grievance mechanisms. Developments regarding the digital platforms in the subregion address content moderation beyond a takedown-centric approach, including a broader set of platform governance tools such as ranking, demotion, demonetization, and account suspension. However, participatory policymaking processes and international standards, such as the UNESCO Guidelines for the Governance of Digital Platforms, are impacting governance debates in the subregion. This toolkit is an outcome of the consultations held as a part of a joint initiative, "Safeguarding Freedom of Expression and Access to Information through the Implementation of the UNESCO Guidelines for the Governance of Digital Platforms" undertaken by Tribhuvan University, Nepal, Jahangirnagar University, Bangladesh and University of Vocational Technology, Sri Lanka and supported by UNESCO and the European Union.

There is an evolving but fragmented regulatory landscape across those three countries. As the governments are seeking to regulate digital platforms, their approaches vary significantly and raise concerns about freedom of expression, institutional independence and the balance between state authority and platform accountability. Bangladesh, Nepal and Sri Lanka all guarantee freedom of expression in their constitutions. Article 39 of the Constitution of Bangladesh, Article 19(2) of the Constitution of Nepal and Article

14(1)(a) of the Constitution of Sri Lanka recognize the right to freedom of expression and publication. However, these rights are subject to certain legal restrictions, creating ongoing debates about the balance between regulation, national security and protection of fundamental freedoms in digital spaces. However, the digital governance landscape in those countries remains in a transitional phase. While the government has attempted to strengthen regulatory oversight of digital platforms, policy development has often been criticized for limited consultation with stakeholders and insufficient safeguards for freedom of expression and the right to access to information.

Bangladesh

Freedom of expression in Bangladesh is constitutionally protected under Article 39 of the Constitution of Bangladesh, which guarantees the right to freedom of thought, speech and press (Government of Bangladesh, 1972). However, this right is subject to regulations imposed by law in the interests of national security, public order, morality or relations with foreign states.

In recent years, Bangladesh has adopted several laws aimed at regulating digital communication and online platforms. The Digital Security Act (2018), widely criticized by journalists, civil society organizations and international human rights groups, was replaced by the Cyber Security Act (2023) which was also criticized as the law could be used to restrict free speech and criminalize legitimate online expression (Government of Bangladesh, 2023).

Following political changes in 2024, the interim government introduced the Cyber Protection Ordinance (2025), repealing the Cyber Security Act. The ordinance was a part of reforms to address concerns raised by civil society and the Gen Z movement and it included recognition of internet access as a civic right. But there are concerns over the provisions that allow surveillance, warrantless search, content blocking and state control over the regulatory body.

Institutionally, the Bangladesh Telecommunication Regulatory Commission (BTRC) plays a central role in regulating digital communications and enforcing platform-related regulations. Under the Cyber Protection Ordinance, the BTRC has the authority to suspend or restrict digital platforms that violate national laws or regulatory requirements (Bangladesh Telecommunication Regulatory Commission [BTRC], 2023).

Nepal

Freedom of expression in Nepal is constitutionally protected, which guarantees freedom of opinion and expression as well as strong protections for media freedom (Government of Nepal, 2015). The constitution explicitly prohibits prior censorship and prevents the closure or seizure of media outlets solely for publishing or broadcasting content. However, the constitution also allows the state to enact laws regulating media and digital communication.

Nepal does not have a comprehensive law specifically governing digital platforms. However, several legal instruments and policy directives have been used to regulate online communication. The Electronic Transactions Act (2008), which was originally intended to address electronic commerce and cybercrime but has often been used in cases involving online speech. The government introduced the Directives for Managing the Use of Social Networks (2023), which required social media platforms to register in Nepal, appoint local contact persons, and establish grievance-handling mechanisms.

The Social Media Bill (2025), which sought to establish a more formal regulatory framework for digital platforms, was withdrawn by the government in January 2026 following public criticism and political changes after the youth-led protests in September 2025. The protests were triggered in part by the government's attempt to block several social media platforms under the 2023 directive (Associated Press, 2025; Reuters, 2025).

Institutionally, the Ministry of Communication and Information Technology (MoCIT) plays a central role in shaping digital platform governance in Nepal. The Nepal Telecommunications Authority (NTA) is responsible for implementing regulatory decisions, including directing internet service providers to enforce platform restrictions. Other institutions, including the Department of Information Technology, Press Council Nepal and the Cyber Bureau of Nepal Police, also address various aspects of online harms, such as misinformation, cybercrime and online harassment.

Sri Lanka

Freedom of expression in Sri Lanka is guaranteed under Article 14(1)(a) of the Constitution of the Democratic Socialist Republic of Sri Lanka, which recognizes the right to freedom of speech and expression, including publication (Government of Sri Lanka, 1978). However, this right is subject to restrictions prescribed by law in the interests of national security, public order and other considerations.

Sri Lanka has recently introduced several laws aimed at strengthening oversight of digital communication and online platforms. The Online Safety Act, No. 9 of 2024, which criminalizes a range of online offences and provides for the establishment of an Online Safety Commission. This commission is intended to oversee harmful online content, investigate violations, and issue directives to digital platforms and internet intermediaries.

Under the Act, the proposed commission would have the authority to order the removal of prohibited statements and direct internet service providers to disable access to specific online locations, including accounts, pages, or websites associated with repeated violations. The law also allows the commission to issue codes of practice and requires digital platforms operating in Sri Lanka to register with regulatory authorities.

In addition to the Online Safety Act, Sri Lanka enacted the Personal Data Protection Act, No. 9 of 2022, which establishes a Data Protection Authority responsible for regulating the processing and protection of personal data.

Together, the OSA and PDPA constitute the backbone of Sri Lanka's current platform governance framework. However, implementation remains incomplete. For example, the Online Safety Commission, envisioned by the Online Safety Act, has yet to be fully operationalized. Recent trends in grievance mechanisms show that the country is increasingly adopting institution-based systems, rather than relying solely on private platform complaint mechanisms. As a result, Sri Lanka's system remains in transition, with ongoing improvements in perceived independence and safeguards (Centre for Policy Alternatives [CPA], n.d.).

Bangladesh, Nepal, and Sri Lanka: Common Concerns

Despite differences in regulatory frameworks, Bangladesh, Nepal, and Sri Lanka face several shared challenges in governing digital platforms.

With high internet penetration, these three countries share notable similarities and concerns. In particular, their democratic systems are becoming increasingly vulnerable to threats to information integrity and cyber harms. Threats related to the information ecosystem have contributed to political unrest, increased violence, distorted public discourse and declining trust in institutions across the region. The digital platforms are expected to play a constructive role in managing above mentioned challenges. Hence, coordinated government communication strategies, stronger platform accountability and sustained media and digital literacy initiatives have become common priorities.

None of these countries currently has a comprehensive framework explicitly regulating algorithmic ranking systems, recommender algorithms, demotion practices, or

transparency reporting obligations in the manner of more advanced regimes such as the European Union's Digital Services Act (DSA). Since the act regulates how digital platforms, online marketplaces, social media companies, and search engines handle illegal content, disinformation, user rights, and platform risks, such approach is relevant in Nepal, Sri Lanka and Bangladesh as well.

Therefore, the state of digital rights and governance of digital platforms remains in a transitional phase in Bangladesh, Nepal, and Sri Lanka characterized by changing regulatory frameworks, institutional uncertainties and ongoing tensions between regulation, freedom of expression and platform accountability. Overall, the region is experiencing a transition toward regulation of digital platforms, but their governance approaches remain fragmented, enforcement-driven and not aligned with international human rights standards.

The following common trends are seen across Bangladesh, Nepal and Sri Lanka.

1. Governments are demonstrating a growing tendency to regulate digital platforms, moving beyond merely discouraging harmful content, as reflected in recent legislation.
2. A significant proportion of the population actively uses digital platforms; however, overall media and digital literacy levels remain low.
3. Regulatory bodies often lack clear mandates, adequate institutional capacity, and well-defined jurisdiction to effectively protect users from harmful content.
4. A communication gap persists between platform operators and regulators, which becomes more pronounced when governments act without adopting a multistakeholder approach.
5. There is an emerging sense of collective effort among governments in the region to address the influence and accountability of large platform companies.
6. The impact and effectiveness of media are declining, while digital platforms are increasingly becoming primary vectors for disinformation, misinformation, hate speech and other harmful content.
7. There is a pressing need for clearer frameworks to content moderation particularly addressing violence against women, the best interests of children, caste-based hate speech targeted to religious and ethnic minorities, the protection of electoral integrity and democratic processes, and the prevention of online harms.
8. The rise of "sponsored" or "boosted" content that mimics organic social interactions, making it harder for users to distinguish between advertising/ propaganda and friend-based content.

Unit I

Supporting Freedom of Expression and Access to Information

Freedom of expression and access to information are foundational principles of democratic societies and essential components of a healthy digital environment. These rights include freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers. These fundamental rights enable individuals to express opinions, share knowledge, participate in public debate and access diverse sources of information. In contemporary societies, online platforms have become central for exercising these rights.

Digital platforms facilitate unprecedented opportunities for participation in public communication and discourse. However, they also create governance challenges, as the platforms not only enable the rapid dissemination of information but can also amplify misinformation, harassment, and harmful content. The digital platforms' interventions to minimize harmful content such as content moderation are often opaque in nature and are vulnerable to manipulation. The government interventions to regulate digital spaces endangers legitimate expression if such interventions are not aligned with international human rights standards.

International human rights frameworks, including Article 19 of the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights (ICCPR), recognize the right to free expression (United Nations, 1948, 1966). The UNESCO Guidelines for the Governance of Digital Platforms emphasize that platform governance must be grounded in these principles while addressing online harms (UNESCO, 2023).

Within this framework, civil society plays an important role in safeguarding freedom of expression and access to information in digital environments. Civil society organizations comprising media institutions, academic researchers, and digital rights advocates can contribute to ensuring that platform governance and platform practices remain transparent, accountable, and aligned with human rights standards. Through research, monitoring, advocacy and public engagement, civil society helps balance the need to address online harms with the protection of these fundamental freedoms.

This unit is designed for media freedom organizations, human rights defenders, legal advocates, right to information advocates, and researchers. It supports those

working to safeguard constitutional rights, such as freedom of expression, privacy and right to information, and to influence policies related to digital communication and platform governance.

What this unit provides?

- **Evidence Collection Framework:** A structured framework to document cases of online censorship through content moderation (arbitrary content removal, account suspension and platform blocking, etc.) enabling their use in legal and advocacy efforts.
- **Regulatory Tracking:** Key indicators to monitor the actions of state agencies and assess how law enforcement practices impact online expression and the right to access to information.
- **Policy Benchmarking:** Offers guidance to evaluate proposed laws and policy against international human rights standards.

What can civil society do?

- Support freedom of expression and access to information through monitoring, research, public engagement, and policy advocacy.
- Understand how platform policies, government regulations, and technologies affect the visibility, virality, and accessibility of online information.
- Monitor digital platforms' content moderation, community guidelines and algorithmic systems.
- Check whether harmful content is being removed fairly and whether bias exists against certain languages, communities or viewpoints.
- Document threats such as unjust content removal, account suspension, platform blocking, and legal pressure on online speech.
- Use evidence to identify patterns, raise awareness, and inform policy debates. Document threats to online expression arising from content moderation practices, including unjust removal of content, account suspensions, platform blocking, and the application of legal or policy-based pressures that may restrict lawful speech.
- Promote access to trustworthy information through media and information literacy, public interest journalism, local media support, and fact-checking partnerships.
- Engage with policymakers and platforms by providing research, joining consultations,

BOX 1: THE 'RED FLAG' LEGISLATIVE AUDIT

When a new digital law is proposed, this quick checklist can be used to evaluate it:

Civil Society Checklist: Assessing Social Media Laws for Risks to Freedom of Expression (FoE) and Freedom of Information (FoI)

- **Vagueness and value-based restrictions:** Does the law use broad, undefined, or subjective terms such as "fake news," "offensive content," "harmful speech," or "moral values" without clear legal definitions and narrow, foreseeable scope? *(If yes: High risk to FoE/FoI)*
- **Criminalization of expression (ultima ratio principle)**: Does the law criminalize forms of expression, and if so, is criminal liability applied only as a last resort, with a clearly defined threshold of harm and consideration of non-criminal alternatives? *(If criminalization is broad or primary tool: High risk)*
- **Proportionality of sanctions:** Are penalties (fines, imprisonment, account restrictions, or platform blocking) proportionate to the severity of harm, or do they risk creating a chilling effect on lawful expression? *(If disproportionate: High risk)*
- **Governance and oversight of restrictions:** Are content removal orders, account suspensions, or platform blocking decisions subject to independent and impartial oversight? Does this include judicial review and/or a genuinely independent regulatory authority with clear due process safeguards? *(If no independent oversight: High risk)*
- **Procedural safeguards and due process**: Are individuals and platforms guaranteed notification, clear reasoning for restrictions, and the ability to challenge decisions through timely and accessible procedures? *(If absent: High risk)*
- **Mandatory identification and anonymity:** Does the law require real-name registration or government-issued identification to access social media services, thereby undermining anonymity and increasing risks for journalists, activists, and vulnerable groups? *(If yes: High risk)*
- **Good Samaritan principle and general monitoring prohibition:** Does the law preserve the "Good Samaritan" principle by protecting platforms that voluntarily moderate content in good faith, and does it explicitly prohibit general monitoring obligations or filtering requirements that would force platforms to systematically monitor all user content? *(If general monitoring is required or Good Samaritan protections are weak: High risk)*

and advocating human-rights-based governance.

- Encourage multistakeholder cooperation among governments, platforms, universities, researchers and technical experts.
- Ensure marginalized and vulnerable communities are represented in digital governance discussions.

Action Points for Civil Society

Civil society organizations can take actions to promote freedom of expression and access to information in digital platform governance. Following the core CSOs accountability cycle, key actions to support freedom of expression include:

1. Monitoring and documentation

a) Monitor digital platform policies and moderation practices regarding online content.
b) Document the incidents of content moderation/curation, account suspension or platform restrictions affecting freedom of expression and access to information.
c) Track government regulations, legislative initiatives and state agencies' practices relevant to digital communication and platform governance.
d) Monitor law enforcement agencies' practices regarding online expression by the citizens (Government of Bangladesh, 2009; Government of Nepal, 2007; Government of Sri Lanka, 2016).
e) Monitor violations of digital rights focusing on surveillance and data privacy by state, private entities and digital platforms.
f) Engage in dialogues regarding user verification processes to reduce the proliferation of anonymous fake accounts while balancing the right to privacy.

2. Research

a) Conduct research on the impact of platform policies on access to information and public discourse.
b) Study algorithmic amplification and how recommendation systems influence information visibility.
c) Conduct research on public perception and users' experience on online communication.
d) Review bills and policies vis-à-vis international practices and human rights standards.

e) Conduct surveys to understand the knowledge and skills in media and information literacy among the public.
f) Produce reports and policy briefs.

3. Advocacy and policy engagement

a) Advocate for platform governance frameworks aligned with international human rights standards.
b) Participate in policy consultations, parliamentary hearings and regulatory processes.
c) Engage with digital platforms to encourage transparency in moderation policies and decision-making processes.
d) Review the state of freedom of expression, harmful content, state policies and digital platforms' practices periodically.

4. Public awareness and digital literacy

a) Promote media and information literacy programs to help users critically evaluate online information.
b) Raise public awareness about digital rights, online safety and responsible digital participation.
c) Support fact-checking initiatives and public interest journalism.
d) Assess needs, develop programs and implement public awareness and skills-development interventions among users on media and information literacy.

5. Multistakeholder collaboration

a) Build partnerships among media organizations, rights groups, universities, research organizations, technology experts, and international networks.
b) Participate in local, national, regional and global digital governance forums.
c) Facilitate dialogue among users, policymakers and digital platforms to improve accountability and transparency. Implementation Checklists, Tools and Monitoring Indicators for Supporting Freedom of Expression & Access to Information

Implementation Checklists, Tools and Monitoring Indicators for Supporting Freedom of Expression & Access to Information

Implementation Checklist	Implementation Methods	Monitoring Indicators
1. Monitor digital platforms for the de-prioritization of critical voices and local news outlets to ensure information remains accessible.	Visibility Audit: Search to verify if specific political keywords, local news topics or activist handles are hidden in "top" feeds.	# of local news outlets and critical voices reporting unexplained reach drops # of verified instances where algorithmic de-prioritization was detected.
2. Internet shutdowns and throttling: Document and monitor instances of internet shutdowns or the intentional throttling (slowing) of social media apps and websites.	Network connectivity log: A daily tracker of upload/ download speeds and latency. Open Observatory of Network Interference (OONI) Probe Tool to collect data and OONI Explore Tool to see collected data.	# of documented incidents and the total duration of restricted or blocked digital access for the public.
3. Policy Audit: Audit national legislation to identify provisions for surveillance, private data access and threats to digital privacy.	Legal Compliance Matrix: A 3-column analysis tool comparing: (1) UNESCO/ ICCPR Principles, (2) Local Law Clauses and (3) Potential rights violations.	# of specific legal clauses identified as restrictive # of documented violation of data privacy and surveillance.
4. Criminalization and prosecution tracking: Monitor and track law enforcement practices regarding the criminalization of online expression and journalism.	Tracker: A database tracking all incidents including data of victims, case summary, provision of law used to build case, and court responses.	# of people arrested or questioned for legitimate online opinions or journalistic work # of cases successfully challenged in court.

Implementation Checklist	Implementation Methods	Monitoring Indicators
5. Platform transparency reporting as a complementary source of evidence: In addition to reviewing Community Guidelines against constitutional and legal standards, it is recommended to systematically draw on platform transparency reports where available.	Legal Alignment Review: A policy assessment tool used to flag where platform-wide rules unfairly restrict speech that is otherwise legally protected under the local laws. Review the state of content moderation practices, volume and categories of content removed, government or legal request received, enforcement trend over time. It is necessary to triangulate whether identified policy–law conflicts are reflected in actual enforcement patterns and scale.	# of specific platform clauses identified as being harmful to local free speech # number of complaints sent to platforms regarding policy-law conflicts. # alongside "number of specific platform clauses identified as harmful to local free speech" #a metric capturing relevant transparency report disclosures (e.g. takedown volumes, government requests, and appeals outcomes) to better assess systemic impact and platform responsiveness.
6. Self-Censorship and Surveillance Trends: Track the prevalence of self-censorship among media professionals and document specific violations of digital rights through targeted surveillance.	Survey: A confidential survey for journalists to report threats, topics they avoid for safety and suspected digital surveillance incidents.	# of journalists admitting to self-censorship for safety # of incidents of surveillance acts targeting human rights defenders.

Unit II

Addressing Harmful Content and Identifying Roots of Abusive Behaviours

Abusive behaviour in digital spaces is a growing challenge affecting users and non-users of digital platforms, including individuals, communities. It includes harassment, hate speech, intimidation, coordinated trolling, disinformation, and exclusionary practices that often target vulnerable groups such as women, minorities, and persons with disabilities.

Such behaviour causes psychological harm, reputational damage, and social exclusion, while also weakening democratic debate and trust in institutions. Although digital platforms enable communication and participation, they can also amplify harm when not properly governed.

The UNESCO Guidelines for the Governance of Digital Platforms stresses a human rights–based approach, including assessing how platform design, algorithms, and content moderation systems impact different users.

Abusive behaviour is driven by broader social and structural factors such as discrimination, prejudice, polarization, information disorder, stereotypes, and low media and information literacy, alongside platform features like algorithmic amplification, and engagement-based recommendations.

Civil society plays a key role in identifying abuse, analyzing root causes, and supporting accountability through monitoring, research, and advocacy for safer digital environments.

What this unit provides?

- **Structural Root Cause Analysis:** Identifies systemic drivers of online abuse, such as platform design features (including ranking and recommendation systems), content amplification dynamics, and underlying societal inequalities, to support long-term responses that go beyond surface-level symptoms.
- **Impact Assessment:** Offers guidance for examining the social and psychological effects of others, harassment, and exclusionary online behaviour.

- **Data Disaggregation:** Encourages collection of data on how abuse affects vulnerable communities to support evidence-based policies and interventions.
- **Strategic Support Guidelines:** Provides step-by-step approaches for establishing helplines, referral services, and secure reporting mechanisms for victims.
- **Classification Tools:** Supports the categorization of different forms of online harm using consistent and precise terminology to enable more targeted and proportionate responses.
- **Monitoring Framework:** Includes tools to track threat actors, targeted groups, abuse patterns, and impacts on public discourse.

What Can Civil Society Do?

- Monitor digital spaces to detect harmful trends, harassment, hate speech, coordinated manipulation, and viral misinformation.
- Identify threat actors and networks that create or amplify abusive content and understand how harmful narratives go viral.
- Document abusive content moderation/curation by collecting evidence on content, targeted groups, scale of amplification, and impacts.
- Analyze root causes such as platform design, weak moderation systems, algorithmic amplification, and social inequalities created through digital divide.
- Engage affected communities through observation, consultations and research to get lived experiences.
- Support victims through legal guidance, psychological assistance, technical help, and referral services to relevant organizations.
- Promote media and information literacy by helping users recognize misinformation, harmful content, and safe online responses.
- Raise public awareness through campaigns, workshops, and community dialogue on digital behaviour.
- Advocate for accountable governance by pushing platforms and policymakers toward transparent, rights-respecting, and culturally sensitive moderation.
- Use research for policy change by translating findings into evidence-based recommendations for governments, platforms, and international stakeholders.

Action Points for Civil Society

Civil society organizations can take following actions to address harmful online content while protecting freedom of expression.

1. Monitoring and documenting the changing behaviours

a) Monitoring the behaviours of specific groups such as children, women, minorities and journalists on different digital platforms and products emerging over the time.
b) Monitor digital platforms to identify harmful content including hate speech, harassment, information disorder, manipulated media and online scams.
c) Document incidents of online abuse, harassment campaigns or coordinated manipulation targeting individuals or communities.
d) Track patterns of harmful narratives, including those targeting women, journalists, human rights defenders and marginalized groups.
e) Maintain databases or records of harmful content cases to support research, reporting and advocacy.

2. Research

a) Conduct research on how misinformation, disinformation and abusive content spread on digital platforms.
b) Analyze how platform design, recommendation algorithm and engagement-driven systems amplify harmful content.
c) Study the impact of harmful content on democratic processes, social cohesion, diversity and vulnerable communities as well as individuals.
d) Publish research reports, policy briefs and evidence-based analyses that inform public debate and policy development.

3. Community reporting and support

a) Establish reporting channels such as helplines, email addresses or secure online forms where individuals can report harmful online content or abuse.
b) Develop clear and standard procedures for receiving, documenting and responding to reports of new risks on online services and products.
c) Build referral networks with legal aid providers, psychosocial counsellors and digital security experts to provide support to victims.
d) Provide guidance to individuals on how to report harmful content to platforms or relevant authorities.

4. Legal and Psychological Assistance

a) Partner with legal aid organizations to assist individuals facing online harassment, defamation or digital rights violations.
b) Provide or facilitate trauma-informed counselling and support services for victims of online abuse.
c) Offer guidance on legal remedies, complaint mechanisms and digital platform reporting procedures.
d) Develop Psychosocial Scaling frameworks to assess the trauma levels of monitors and victims, ensuring that support is tailored to the severity of the online abuse and the specific low-trust environment.

5. Digital literacy and public awareness

a) Conduct digital literacy and media and information literacy activities to help communities recognize misinformation and harmful content online.
b) Raise awareness on online participation and digital communication to strengthen the concept of digital citizens in order to create community led codes of conduct for online groups and civic spaces (https://unesdoc.unesco.org/ark:/48223/pf0000227729).
c) Educate users on digital rights, privacy protection and safe online practices.
d) Collaborate with educators, journalists and community organizations to promote media and information literacy (World Health Organization [WHO], 2020). For more information, consider the action of the MIL action plan: *https://www.unesco.org/en/articles/unesco-launches-multistakeholder-action-plan-empower-users-digital-platforms*

6. Creating safe online spaces

a) Facilitate moderated online communities where individuals can safely discuss online harms and share experiences.
b) Establishing clear codes of conduct and moderation guidelines to ensure trustworthy and safe participation.
c) Use insights from community discussions to identify emerging patterns of online abuse and inform advocacy efforts.
d) Develop psychosocial scaling frameworks tailored to low-trust settings, ensuring that trauma-informed care is culturally appropriate and locally accessible.

7. Engagement with digital platforms

a) Engage with digital platforms to encourage transparent, fair and rights-respecting content moderation practices.
b) Advocate for stronger reporting systems, user protection mechanisms and transparency in moderation decisions.
c) Encourage publications of due diligence practices and privacy relevant information.
d) Provide feedback to platforms on language-specific moderation challenges and national or regional content issues.
e) Develop the risks assessment measures during elections, protest or emergencies that allow platforms to identify the mitigation measure they have to put in place in a given context.

8. Multistakeholder engagement

- Advocate for balanced regulatory frameworks that address harmful content and protect freedom of expression.
- Participate in policy consultations, public hearings and multistakeholder discussions on digital platform governance.
- Collaborate with media organizations, academic institutions, fact-checking institutions and digital rights groups to strengthen due diligence and accountability.
- Contribute to national and regional initiatives aimed at improving the safety and integrity of digital information ecosystems.

BOX 1: HOW TO ENGAGE AGAINST HARMFUL ONLINE CONTENT

Civil society organizations play a vital role in protecting digital rights by addressing harmful online content, supporting affected communities and strengthening safe online environments. Effective engagement requires structured approaches that combine immediate support, legal guidance, counselling and community-based platforms. Below are three core strategies elaborated step by step for practical implementation.

1. Establish helplines and referral services

a) Setting up helplines, email addresses or secure online reporting forms.
b) Developing clear procedures for receiving and responding to reports.
c) Building referral networks with legal aid providers, counsellors, digital security experts and relevant authorities.
d) Collecting anonymized data on reported cases to identify trends.
e) Ensuring services are accessible to diverse communities, including marginalized groups and speakers of different languages.
f) Establish "Monitor Support Systems" to provide trauma-informed care for civil society workers who are exposed to graphic/harmful content during the monitoring process (SAMHSA, 2014).

2. Provide counselling and legal assistance

Victims of online harassment, disinformation campaigns or digital threats require legal, psychological or technical support.

Civil society organizations can:

a) Partner with legal aid organizations to provide guidance on cybercrime complaints, defamation cases and takedown requests.
b) Offer counselling or referral services.
c) Conduct training workshops on digital safety, privacy protection and responding to online abuse.
d) Establish rapid response mechanisms to support individuals targeted by coordinated harassment campaigns.
e) Document cases to identify patterns.

3. Create safe online spaces for marginalized groups

Civil society organizations can facilitate moderated online communities where individuals can safely discuss digital harms and share experiences.

a) Creating moderated online forums or discussion groups on secure platforms.
b) Establishing clear codes of conduct and privacy guidelines to ensure respectful engagement.
c) Using discussions to understand insights into online harms.
d) Supporting advocacy initiatives that support affected communities.

BOX 2: MONITORING FRAMEWORK FOR HARMFUL CONTENT

Monitoring Area	Key Questions	Indicators	Monitoring Methods
Type of Harmful Content	What type of harmful content is circulating?	Hate speech, harassment, misinformation, disinformation, incitement to violence, scams, manipulated media	Social media monitoring, keyword tracking, fact-checking
Targeted Groups	Who is being targeted by harmful content?	Women, journalists, ethnic or religious minorities, activists, political actors	Content analysis, community reporting, victim interviews
Threat Actors	Who is producing or amplifying the content?	Individual accounts, coordinated networks, political actors, anonymous groups	Network analysis, account behaviour analysis
Platform Amplification	How widely is the content spreading?	Shares, likes, comments, trending hashtags, algorithmic amplification	Platform analytics, engagement metrics
Impact on Public Discourse	What are the social or political effects of the content?	Increased polarization, harassment campaigns, misinformation influencing public debate	Media monitoring, public sentiment analysis
Platform Response	How are platforms responding to harmful content?	Content removal, labeling, account suspension, limited action	Platform reporting tools, transparency reports
Legal Context	Are there laws or policies affecting responses to harmful content?	Government directives, regulatory actions, platform policy changes	Legal review, policy tracking

Implementation Checklists, Tools and Monitoring Indicators for Harmful Content and Behaviours Online

Implementation Checklist	Implementation Tools	Monitoring Indicators
1. Helpline and Victim Reporting: Set up and maintain secure, accessible reporting channels (such as helplines, email or encrypted web forms) to report cases of online scams, harassment and digital harm.	**Intake Form:** A digital form with fields for: (1) Type of harm, (2) Platform used, (3) Evidence links/screenshots, (4) Personal impact, and (5) Urgency level (High/Low).	# of reports received.
2. Content Classification: Accurately categorize flagged content based on the intent of the sender and the scale of the harm.	**Harm Taxonomy Guide:** A quick reference with key terminology and their characteristics clearly defined.	# of harmful posts categorized # accuracy in labeling reported content.
3. Support Directory and Referral Network: Map professional service providers and develop a directory that support organizations and victims can use for necessary legal and psychological assistance.	**Referral Directory:** A live database of professional service providers with services offered contacts and costs.	# of victims referred to professional support services # of victims connected receiving help.
4. Safe Online Spaces: Facilitate moderated online communities and forums where marginalized groups can safely discuss digital harm and share lived experiences without fear of further abuse.	**Moderation Playbook:** A set of internal rules for vetting community members, managing sensitive triggers, and preventing re-traumatization in safe space discussion groups.	# of participants in CSO-led safe online communities # of moderated safe spaces established.
5. Deepfake and AI Harassment Tracking: Monitor and document cases of Non-Consensual Intimate Imagery (NCII) or other harassment generated by Artificial Intelligence.	**Harassment Tracker:** A specialized log designed to record AI-generated harassment cases, documenting the scale of amplification and the specific groups targeted.	# of AI-generated harassment cases documented # of AI-generated reports submitted to platforms for action.

Implementation Checklist	Implementation Tools	Monitoring Indicators
6. Identity-Based Narrative Tracking: Monitor and document specific, recurring harmful narratives and hate speech.	**Narrative Audit Tool:** A template used to identify recurring keywords.	# of recurring hate narratives identified.
7. Media and Information Literacy Campaign: Launch awareness campaigns to help users identify, respond and report online harms.	**Digital Literacy Kit:** A toolkit on definitions, identification, responses and reporting mechanisms.	# Engagement of digital literacy materials shared online # participants of campaigns

Action Points for Civil Society

Civil society organizations can take the following actions to identify abusive behaviour and address its root causes within digital environments.

1. Monitoring and detection

a) Monitor digital platforms to identify patterns of harassment, hate speech, coordinated abuse and targeted misinformation campaigns.
b) Track online narratives that incite hostility or discrimination against specific communities.
c) Identify emerging trends in abusive behaviour across different platforms and digital spaces.

2. Data collection and documentation

a) Collect and document evidence of abusive online behaviour, including screenshots, links and engagement metrics.
b) Maintain databases of incidents to track patterns over time.
c) Gather disaggregated data to understand how abuse affects different groups, including women, journalists, minorities and activists.
d) Publish research reports and policy briefs on patterns of abusive behaviour and risks within digital platforms.

BOX 3: UNDERSTANDING ABUSIVE BEHAVIOUR

UNICEF and UN Women's definition of abusive behaviour equally fits in the case of online engagement. According to the definition, abusive behaviour can be listed as a pattern of behaviour in any relationship that is used to gain or maintain power and control. Furthermore, such behaviour could be physical, emotional, sexual, economic or psychological actions or threats of actions that influence another person. UNICEF categorically presents forms of online abuse/violence such as harassment and stalking, image-based abuse non-consensual sharing, doxing, hate speech and trolling, sextortion and online grooming (UNICEF, 2026; UN Women, 2025).

Common Forms of Abusive Behaviour

- Harassment: Repeated or targeted attacks on individuals or groups.
- Hate speech: Content expressing hostility based on group identity race, ethnicity, gender, religion, disability or other protected characteristics.
- Threats and intimidation: Messages or actions intended to create fear of harm.
- Disinformation and manipulative content: False or misleading information designed to deceive, harm or influence public opinion against individuals or groups.
- Exclusionary practices: Online actions that systematically marginalize, silence or intimidate voices in digital spaces.

Why the Severity of Abusive Behaviour Matters?

- Abusive behaviour can have significant consequences for individuals and communities:
- Psychological and emotional harm: Exposure to harassment or threats may lead to stress, anxiety, depression or trauma.
- Social exclusion: Targeted abuse may silence individuals or communities, limiting their participation in public discourse.
- Reputational damage: Individuals and organizations suffer harm to their credibility, employment prospects or public trust.
- Reinforcement of systemic inequality: Marginalized groups often experience disproportionate levels of online abuse, reinforcing existing social inequalities and stereotypes.

BOX 4: ROOT CAUSES OF ABUSIVE BEHAVIOUR

Understanding the root causes of abusive behaviour is essential for designing effective prevention, policy and platform governance measures. Online abuse is often driven by a combination of structural, technological, social and political factors.

Key root causes include:

- Pre-existing social inequalities: Discrimination based on gender, race, ethnicity, religion, disability, sexual orientation or socioeconomic status often translates into digital spaces. In the South Asian context, these inequalities are often deepened by Colonial Legacies that institutionalized certain ethnic and religious divides, which are now weaponized through digital "othering"; and exclusionary online practices.
- In South Asia, these divides are frequently exacerbated by historical and colonial legacies that continue to shape ethnic, caste-based, and religious fault lines online.
- Anonymity and lack of accountability: Anonymous or pseudonymous identities can reduce social accountability and encourage harmful behaviour.
- Algorithmic amplification: Engagement-driven algorithms may unintentionally prioritize sensational, polarizing, or harmful content, increasing the visibility of abusive behaviour.
- Platform design and features: Certain platform functionalities such as viral sharing, comment systems, or weak reporting tools can facilitate the spread of harassment or coordinated abuse.
- Political polarization: Highly polarized political environments can fuel hostile online discourse and targeted harassment campaigns.
- Misinformation and disinformation ecosystems: False or misleading information can trigger hostility, scapegoating and attacks against individuals or groups.
- Digital literacy gaps: Limited understanding of responsible online behaviour, media and information literacy, and platform policies can contribute to harmful interactions.
- Socio-Cultural norms: In some contexts, harmful stereotypes or normalized discriminatory attitudes may shape online behaviour.
- Weak governance and enforcement: Inadequate platform moderation, inconsistent policy enforcement, or lack of regulatory oversight can allow abusive behaviour to persist.
- Economic incentives of platforms: Business models that prioritize engagement and attention may indirectly incentivize controversial or harmful content.

- Power imbalances and silencing tactics: Abusive behaviour may be used strategically to intimidate or silence journalists, activists, human rights defenders, and marginalized voices.
- Psychological and behavioural factors: Online disinhibition, mob behaviour, and group dynamics can encourage individuals to engage in actions they might avoid offline.

Identifying abusive behaviour is vital to reducing online harm. Recognizing and documenting abusive behaviour enables effective responses and strengthens digital governance. Early detection allows regulators and platforms to act before abuse escalates, helping to prevent further harm. It is important for protecting women and other vulnerable groups and marginalized communities including ethnic minorities, and persons with disabilities who are often disproportionately targeted. Evidence of abusive behaviour also helps improve platform governance by refining moderation policies and strengthening transparency. In addition, monitoring abuse promotes accountability and human rights by enabling oversight of platform practices. Data on abuse patterns supports policy and advocacy efforts, helping civil society push for stronger protections and safeguards. Identifying abusive behaviour raises public awareness by helping users recognize harmful content and respond responsibly.

These steps enable researchers and policymakers to address not only individual incidents of abuse but also systemic factors that enable harmful behaviour to emerge and spread in digital environments.

3. Research and analysis

a) Conduct research on the root causes of abusive behaviour, including political polarization, information pollution ecosystems and discriminatory social norms.
b) Examine how platform design features such as recommendation algorithms or engagement metrics or privacy settings contribute to the spread of abusive content.
c) Study the social and psychological impacts of online abuse on individuals and communities.

4. Community engagement

a) Engage affected communities to understand their experiences of online abuse.
b) Conduct interviews, surveys or focus group discussions to collect information

from impacted individuals and groups.

c) Help platforms and policymakers to understand local linguistic, cultural and social contexts.

5. Advocacy for transparency and accountability

a) Advocate for increased transparency in platform moderation practices and moderation and curation practices.

b) Encourage platforms to conduct human rights impact assessments and strengthen user protection mechanisms.

c) Advocate for granular transparency reports.

6. Multistakeholder collaboration

d) Build collaboration among civil society organizations and non-profit organizations, academic institutions, think tanks, media organizations, legal groups and digital rights groups to conduct joint research on online abuse.

e) Participate in multistakeholder forums on digital governance and platform accountability.

f) Share research findings with policymakers and international organizations to support evidence-based governance approaches.

Implementation Checklists, Tools and Monitoring Indicators of Identification for Abusive Behaviour and Root Causes

Implementation Checklist	Implementation Tools	Monitoring Indicators
1. Narrative Tracking: Map identity-based hate narratives and monitor harassment campaigns across different digital platforms.	Narrative Mapping Tool: A visual diagram showing: (1) Origin of a lie, (2) Main amplifiers such as bots or influencers, (3) The specific target, and (4) The cross-platform path.	# of hate narratives identified # of mulit-platform abuse campaigns documented.
2. Data Disaggregation and Lived Experiences: Gather specific data on how abuse impacts different groups and document the severity of harm through qualitative insights.	Victim Impact Database & Interview Guide: A structured database for tracking victim demographics paired with a interview guide on psychological and reputational damage.	# of cases recorded. # of in-depth lived experience case studies.
3. Algorithmic Design Analysis: Audit recommendation systems to identify how platform design features contribute to the spread of abuse.	Algorithmic & Design Audit: A log showing when neutral searches lead to harmful content.	# of instances where harmful content appeared in "recommended" feeds # of platform features contributing to increased abuse.
4. Root Cause Mapping: Analyze political polarization and document how social inequalities or structural dynamics drive harmful online behaviour.	Polarization Scorecard & Root Cause Audit: Log of politically polarized narratives and a survey questionnaire to understand root causes.	# of narratives trending during major events # of research papers identifying systemic, non-technical causes of online harm.
5. Evidence-Based Policy Advocacy: Translate research findings into specific, data-backed policy recommendations.	Policy Briefs Template: A template for professional policy briefs.	# of policy briefs. # of research-backed recommendations adopted into digital rights discussions.

Unit III

Fact-Checking and Watchdogging Digital Platforms

In the digital world, media organizations, fact-checking initiatives, civil society groups and media professionals including social media content creators play a critical role in ensuring that the information in the digital ecosystem remains accurate, transparent and accountable. Digital platforms have transformed how news is disseminated but they also carry risks such as misinformation, disinformation and the amplification of harmful content. Fact-checking and watchdog activities help safeguard public trust, uphold freedom of expression and support access to trustworthy information.

Fact-checking and watchdog activities are therefore essential components of platform accountability. Fact-checking refers to the systematic verification of public claims, online narratives and viral content using credible evidence and reliable sources (Africa Check, n.d.) Watchdogging refers to independent monitoring of how digital platforms operate, including their content moderation practices, algorithmic amplification systems, advertising transparency and responses to harmful content.

Media organizations, content creators and journalists are key actors in fact-checking and watchdogging. They play a role in shaping public understanding of social and political issues as they produce and disseminate information for public consumption. Media organizations help expose misinformation, threat actors, counter manipulative narratives and promote evidence-based public discourse through investigative reporting, fact-checking and editorial oversight.

Understanding the different forms of misleading information is essential for effective monitoring. Misinformation refers to any false or misleading information without intention shared to deceive while disinformation is deliberately fabricated or manipulated information intended to mislead audiences. UN concept outlines that misinformation is false, but not created with the intention of causing harm (UNDP, n.d.) Information disorder uses correct information used in harmful contexts, such as exposing private information or presenting facts without context to cause harm.

Information disorder erodes trust in democratic institutions and processes and increases social tensions while also causing psychological and real world harm on individuals. Engagement-driven algorithms on digital platforms prioritize attention grabbing contents that inadvertently spread misleading content.

Fact-checking discourages the spread of misinformation and disinformation, promotes public trust in reliable information, improves transparency and accountability in digital information ecosystems and supports informed public debate.

The UNESCO Guidelines for the Governance of Digital Platforms (2023) emphasize that transparency, accountability and access to information are fundamental to healthy digital ecosystems. Media organizations, journalists, content creators and civil society organizations contribute to these principles by spreading correct information, verifying information, exposing misinformation and monitoring platform governance practices. Such activities help strengthen information integrity, promote responsible digital communication, strengthen journalism and encourage digital platforms to operate with democratic values and human rights.

This unit is suitable for journalists, media, civil society organizations with fact-checking and watchdogging initiatives and information integrity researchers. It focuses on ensuring information integrity and performing independent monitoring of platform operations.

What this unit provides:

- **Standardized Methodology:** A fact-checking methodology from identifying a claim to communicating findings.
- **Narrative Tracking:** Supports journalists in identifying false narratives through a specific typology of misleading information.
- **Watchdog Framework:** Provides action points for investigative journalism into how platforms prioritize attention-grabbing content via algorithms.

What can civil society do?

Civil society can strengthen digital information integrity by:

- Monitoring information ecosystems to detect information disorder, manipulative narratives and coordinated influence campaigns.
- Verifying public claims and viral content using reliable sources, including official records, expert analysis and peer-reviewed research.
- Publishing fact-checks in clear and accessible formats, explaining the evidence and verification methods used.

- Conducting investigative reporting on disinformation, misinformation and information manipulation.
- Examining how digital platforms moderate content, enforce guidelines and respond to harmful or political content.
- Sharing data, tools and findings with journalists, fact-checkers, researchers, academia and digital rights groups.
- Coordinating investigations and amplify verified information to reduce the spread of misleading content.
- Presenting findings in formats that are useful for the public, policymakers and platforms.

Action Points for Civil Society

Civil society organizations, media institutions, journalists, academia and content creators, and fact-checking initiatives can take the following actions to strengthen fact-checking and digital platform accountability.

1. Monitoring

a) Track viral narratives, misleading claims and coordinated disinformation campaigns on digital platforms.
b) Monitor how information spreads across platforms such as the role of influencers, political actors and automated networks.
c) Identify threat actors, contents, strategies and vectors.
d) Identify trends of misleading misinformation trends during sensitive periods such as elections, conflicts, natural disasters and crises.

2. Fact-checking and verification

a) Integrate verification and fact-checking processes using reliable and verifiable sources during content production.
b) Cross-check claims against official documents, expert analysis, public interest journalism and academic research (Journalism Authority, n.d.).
c) Establish fact-checking unit or initiative adhering to the code of principles of the International Fact-Checking Network.
d) Publish fact-checks explaining verification methods and evidence used in the analysis.

e) Collaborate with national and international fact-checking networks to strengthen verification.

3. Investigative journalism

a) Conduct investigative reporting on threat actors of misinformation, disinformation and mal information and coordinated influence campaigns.
b) Examine how digital platforms moderate content and enforce their community guidelines.
c) Investigate the transparency of platform policies, advertising systems and algorithmic recommendations.
d) Report on platform responses to misinformation, disinformation and mal information and harmful content.

4. Research

a) Conduct research on information pollution ecosystem and amplification patterns.
b) Analyze the role of algorithms, engagement metrics, and platform design in expanding information disorder.
c) Produce reports and policy briefs identifying risks to information integrity.

5. Public awareness

a) Pre-bunk potential information disorder based on the patterns and trends during political and other crises.
b) Encourage explanatory journalism in order to help the public to understand risks of information disorder.
c) Promote media and information literacy programs.
d) Partner with educational institutions and community organizations to implement media and information literacy initiatives.

6. Engagement with digital platforms

a) Communicate research findings and fact-checking evidence to digital platforms.
b) Advocate for improved content moderation transparency and stronger reporting mechanisms.
c) Encourage platforms to support fact-checking initiatives and improve responses to information disorders.

7. Advocacy

a) Advocate for platform governance frameworks that promote transparency and accountability while protecting freedom of expression.
b) Participate in multistakeholder dialogues involving governments, media, civil society and technology companies.
c) Contribute expertise to policy consultations on digital governance, information integrity and platform accountability.
d) Media & information Literacy, Digital Literacy now encompass prebunking. Prebunking involves proactively exposing people to common misinformation techniques and manipulation tactics in a controlled educational context, thereby building resilience against future exposure to misleading content. The emphasis should be on strengthening critical thinking, source evaluation, and awareness of emotional manipulation, rather than on disseminating false content itself. It may also be useful to clarify that prebunking has become an increasingly important component of media and information literacy and digital literacy initiatives aimed at fostering informed and resilient online participation.

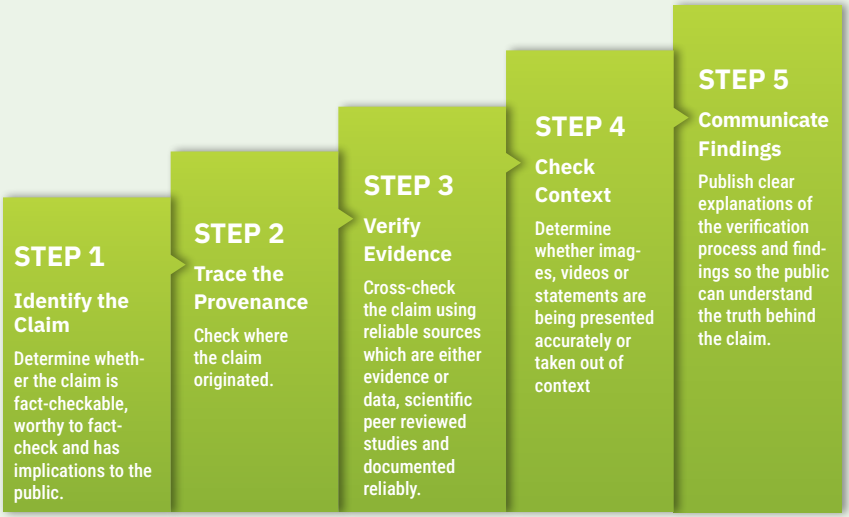
BOX 1: TYPES OF MISLEADING INFORMATION

Understanding these forms of misleading information helps strengthen fact-checking, monitoring, and watchdog efforts, supporting more accountable and trustworthy digital information ecosystems.

- False or fabricated content: Completely untrue information presented as fact.
- Misleading or distorted content: Information framed in a way that creates a false or inaccurate impression.
- Manipulated or synthetic content: Altered images, videos, or audio intended to deceive, including deepfakes.
- Imposter content: Information falsely attributed to credible institutions, experts, or media outlets.
- Satire or parody misinterpreted as fact: Humorous or satirical content mistakenly believed to be true.
- Out-of-context information: Genuine content shared without proper context, leading to misunderstanding.
- Clickbait or sensational content: Exaggerated or misleading headlines designed to attract attention and engagement.

(Adopted from Wardle, 2019

BOX 2: STEPS IN FACT-CHECKING



(Adopted from The Truth in Journalism Fact-Checking Guide, Journalism Authority, Africacheck.org and Macdonald-Kelce Library, University of Tampa)

IMPLEMENTATION CHECKLISTS, TOOLS AND MONITORING INDICATORS OF UNIT IV: FACT-CHECKING & WATCHDOGGING

Implementation Checklist	The Implementation Tool	Monitoring Indicators
1. Verification Standards and Media Integration: Follow professional verification steps and support local media organizations in adopting formal fact-checking workflows for content production.	Verification Workflow Checklist: A structured guide based on: (1) Identify Claim, (2) Trace the Proverance, (3) Verify Evidence, (4) Check Context, and (5) Communicate Findings.	# of verified fact-checks published # of media houses adopting fact-checking workflows.
2. Political Advertising and Funding Audit: Audit the platform's "Ad Library" to ensure transparency in political campaigning, financial disclosures, and targeting parameters (Democracy Reporting International [DRI], n.d.).	Ad Transparency Tracker: A logging tool with fields for: Advertiser identity, Amount spent, Targeting criteria (age/location/religion), and "Paid for by" clarity.	# of political ads found with hidden or foreign funding; and # of ads flagged for non-transparent funding sources.

Implementation Checklist	The Implementation Tool	Monitoring Indicators
3. Monitoring: Maintain an active watchlist of public figures, influencers, and political actors who frequently disseminate unverified or harmful content.	Influencer Audit Log: A database tracking high-profile accounts to document their history of misinformation. Potential Misinformation Database: Record statements with links.	# of accounts monitored # of influencers corrected by fact-checkers.
4. Synthetic Media Identification: Investigate and debunk deepfakes and other AI-generated audio or video content.	Synthetic Media Audit Guide: A technical checklist for identifying visual and audio inconsistencies in AI-generated media.	# of AI-generated cases recorded # of deepfake reports submitted to platforms for removal.
5. Pre-bunking: Conduct pre-bunking campaigns for common misinformation during events such as protests, elections and natural disasters.	Common Trends of information distortion: A table of common misinformation spread during events of crisis.	# of common trends of information distortion # of pre-bunking publications
6. Regional Collaboration: Collaborate with South Asian fact-checkers to share expertise, knowledge and experience (Acharya, 2023).	Collaboration Guide: A joint dashboard sharing real-time data with regional partners.	# of joint cross-border fact-checks published # sharing sessions held.

Unit IV

Promoting Transparency and Accountability of Digital Platforms

Transparency and accountability are essential principles for a human rights based approach on digital platforms governance. In the digital ecosystem where digital platforms have increasingly impactful role on how information is distributed, moderated and amplified, their policies and decisions have significant implications for freedom of expression, access to information and democratic participation.

Digital platforms make decisions such as how contents are ranked, recommended, removed or demoted that affect millions of users. Many times, such decisions are automated by algorithmic systems based on their content moderation policies, advertising models and community guidelines. Other times, such decisions are made by human moderators. But in any case, the process on such decision-making is usually opaque for business reasons which makes it difficult for users, regulators and civil society organizations to understand how platforms operate.

The UNESCO Guidelines for the Governance of Digital Platforms (2023) emphasize that platforms should adopt transparent policies and provide meaningful information about their governance systems. Transparency includes publishing clear community standards, being accountable on moderation activities, disclosing political advertising practices and providing data that allows academia, researchers and civil society to analyze their practices and the issues.

Accountability is the mechanisms that ensure platforms are responsible for their actions and decisions such as user appeal mechanisms, independent oversight bodies, transparency reporting and cooperation with researchers and civil society organizations.

Transparency and accountability are important for ensuring that platform governance systems respect human rights, protect users from harm and operate in the public interest. Monitoring and public analysis by civil society help identify shortcomings in platform governance and support improvements in transparency and user protection.

This unit is suitable for tech policy experts, digital rights advocates, policy advocacy organizations and digital literacy campaigners. It emphasizes examining platform moderation policies, algorithms and transparency practices while also strengthening user awareness and digital literacy (OHCHR, 2020).

What this unit provides:

Transparency Audit: Defines exactly what platform transparency reports should include such as algorithmic ranking data and political advertising libraries.

Educational Curriculum: Provides a 5-step structured guide to "Building Digital Awareness" for young users, including assessment of needs and curriculum integration.

Gap Identification: Helps experts find shortcomings in automated decision-making and advocate for independent audits of platform systems.

What can civil society do?

CSOs, including academia, research groups, rights organizations and media institutions, can improve platform transparency and accountability through research, monitoring, reporting, and advocacy.

- They should understand how platform governance systems are designed and how they affect users and information ecosystems.
- They should oversee the implementation of due diligence process by platforms. To participate in the definition of risks and in the evaluation of the mitigation measures.
- Analyze platform policies and transparency reports on content moderation, harmful content removal, and misinformation response.
- Assess whether platforms provide adequate information about enforcement actions and moderation practices.
- Monitor recommendation systems and algorithms that shape content visibility and virality.
- Study whether these systems amplify harmful or misleading content.
- Advocate for stronger user appeals systems, independent audits and public disclosure of moderation decisions.
- Provide research and recommendations for governance frameworks that ensure transparency while protecting freedom of expression.
- Promote collaboration among journalists, researchers, and civil society groups.
- Use investigative reporting and research publications to improve public understanding of how platforms influence public discourse.

Action Points for Civil Society

Civil society organizations, media institutions, researchers, and journalists can take the following actions to promote transparency and accountability in digital platform governance.

1. Monitoring platform policies and practices

a) Review platform community guidelines and moderation policies to analyze their clarity and consistency.
b) Monitor how platforms enforce their rules on content removal, account suspension and demotion practices.
c) Study how platforms address harmful content, misinformation/disinformation, hate speech and coordinated manipulation campaigns.

2. Analyzing transparency reports

a) Analyze transparency reports published by digital platforms regarding content moderation and enforcement actions.
b) Encourage digital platforms to publish detailed transparency reports.
c) Compare platform reporting practices to assess whether they provide sufficient information for independent oversight.
d) Identify gaps in reporting, particularly regarding algorithmic moderation, political advertising and automated decision-making.

3. Research

a) Conduct investigative reporting on platform governance practices.
b) Study how platform design choices influence the spread of misinformation and harmful content.
c) Publish research findings that help the public understand how digital platforms impact information environments.

4. Advocating for user rights

a) Advocate for stronger user appeals mechanisms when content is removed or accounts are suspended.
b) Promote transparency in political advertising, sponsored content and influencer marketing.

c) Encourage platforms to conduct human rights impact assessments of their policies and systems.

5. Engagement with digital platforms

a) Engage with platform representatives to communicate research findings and user concerns.
b) Provide feedback on platform policies and encourage improvements in moderation transparency.
c) Encourage platforms to support independent research and civil society monitoring.

6. Multistakeholder collaboration

a) Participate in policy consultations on digital platform governance.
b) Advocate for governance frameworks that promote transparency, due diligence, accountability and human rights protection.
c) Collaborate with policymakers, researchers and international organizations to strengthen oversight mechanisms.

BOX 1: WHAT PLATFORM TRANSPARENCY SHOULD INCLUDE

Key elements of platform transparency should include:

- Content moderation policies: Clear explanations of community standards and rules governing acceptable content such as community guidelines.
- Moderation practices and enforcement data: Public reports on the number and types of content removals, account suspensions, and enforcement actions such as transparency report.
- Algorithmic transparency: General information about how recommendation systems, ranking algorithms and automated moderation systems influence the visibility of content.
- Political advertising transparency: Disclosure of political advertisements, sponsors, funding sources and targeting practices such as publicly accessible advertisements library.
- User reporting and appeals mechanisms: Clear processes allowing users to report harmful content and appeal moderation decisions.
- Cooperation with independent researchers: Access to data and information that allows researchers and civil society organizations to study platform impacts responsibly such as API access.

BOX 2: STEPS TO BUILD DIGITAL AWARENESS

Developing digital awareness in young users involves a series of structured steps:

Step 1: Assessment of Learning Needs

Identify the level of digital literacy, risks, and knowledge gaps among learners.

Evaluate access to digital tools and resources to ensure equitable learning opportunities.

Step 2: Curriculum and Syllabus Integration

Incorporate modules on digital literacy, media and Information literacy, critical thinking, and safe online behaviour.

Embed Potter's three building blocks (Skills, Knowledge Structures, Personal Locus) and seven key skills into learning objectives (Potter, 2022).

Align content with age-appropriate learning goals and evolving digital trends

Include training on technical verification tools such as reverse image search and metadata analysis to empower rural users to identify doctored content.

Step 3: Training and Capacity Building

Conduct workshops, seminars, and interactive sessions led by professors, trainers, and media educators.

Include caregivers and mentors in awareness programs to reinforce learning at home. As a global leader in promoting a safe internet, UNESCO can introduce a parental guidance resource within its Media and Information Literacy (MIL) programme.

Mobilization of students and youth.

Step 4: Practical Application and Engagement

Encourage learners to practice critical evaluation of content, fact-checking, and human rights respecting online communication.

Use experiential learning through projects, simulations, and collaborative online activities.

Apply Potter's seven skills in exercises such as media analysis, synthesis of multiple sources, and scenario-based problem-solving.

Step 5: Continuous Evaluation and Feedback

Monitor learners' progress and adapt teaching methods accordingly.

Collect feedback from students, educators, and caregivers to improve program effectiveness.

BOX 3: SEVEN SKILLS OF MEDIA AND INFORMATION LITERACY

Skill	Definition
1. Analysis	Breaking down a message into meaningful elements
2. Evaluation	Judging the value of an element; the judgment is made by comparing a message element to some standard
3. Grouping	Determining which elements are alike in some way; determining how a group of elements are different from other groups of elements
4. Induction	Inferring a pattern across a small set of elements, then generalizing the pattern to all elements in the set
5. Deduction	Using general principles to explain particulars
6. Synthesis	Assembling elements into a new structure
7. Abstraction	Creating a brief, clear, and accurate description capturing the essence of a message in a smaller number of words than the message itself

(Source: Potter, 2021)

BOX 4: STEPS TO BUILD DIGITAL AWARENESS

Tool: 60-Minute 'Spotting the Black Box' Workshop

0–10 min (The Hook):
- Start with a question: *Why do you see different ads or posts than your friends?*
- Introduce the idea of the "black box" algorithm - how platforms decide what to show you.

10–30 min (The Audit):
- Ask students to open their TikTok or Instagram 'For You' feed.
- Have them find at least one post they did not actively choose to follow and reflect on why it appeared.

30–50 min (Critical Thinking):
- Guide students to analyze the post using simple questions inspired by Potter's skills:
What is the purpose of this post?
-Is it trying to sell, influence, entertain, or inform?

50–60 min (Take Action):
- Show students how to take control— how to report, hide, or give feedback on content or ads.
- Introduce the idea of a 'right to appeal' and human rights respecting platform use.

Implementation Checklists, Tools and Monitoring Indicators for Transparency & Accountability

Implementation Checklist	The Implementation Tools	Monitoring Indicators
1. Transparency Report Audit: Audit platform's transparency reports for data gaps and demand that companies publish country-specific Human Rights Impact Assessments.	Transparency Scorecard: A "Yes/No" checklist to evaluate content moderation policies, removed data, algorithmic transparency and advertising practices.	# of data gaps identified for the region Availability of country-specific human rights impact assessment reports (Yes/No).
2. Digital Literacy: Implement and measure digital literacy programs aligning with James Potter's 7 Skills.	Assessment Tests: A pre-and-post participation quiz based to map the change in digital literacy knowledge and skills.	# of increase in literacy scores
3. Influencer Disclosure Assessment: Scrutinize influencers' contents for paid advertisements or sponsored contents if financial arrangements are disclosed.	Sponsored Content Tracker: A database tracking influencer posts to identify paid partnerships that lack disclosures.	# of local influencers violating ad-disclosure rules # of sponsored content without disclosure.
4. Automated Decision-Making Monitoring: Audit automated decision-making to identify errors where machines fail to understand local language and context	Case Log: A log documenting specific instances where automated moderation failed to understand local language and context.	# of cases recorded # of reports to digital platforms.
5. Ad Library: Audit advertisements library to ensure funding sources, especially of political advertising, are transparent.	Ad Transparency Audit Tool: A tracker for identifying sponsor, total spend and targeting parameters within the platform's ad library.	# of advertisements with non-transparent or foreign funding sources.
6. Terms of Service Accessibility: Review platform's Terms of Service to ensure they are available in local languages and dialects. Ensure that not only the "Terms of Service" but also the "User Grievance/Reporting Forms" are available in local dialects and simplified "plain language."	Plain-Language Translation Tool: A tool used to translate and simplify complex legal terms into easy-to-read guides in local language.	Availability of terms of services in local language (yes/no) # of local languages/ dialects fully supported in platform Terms of Service documents.

Unit V

User Grievance Mechanisms and Independent Oversight

It is important to ensure effective grievance mechanisms and oversight systems for accountable digital platform governance. The individual user and user communities must have accessible mechanisms to report harmful content, contest moderation decisions and seek remedies when their rights are affected.

Digital platforms routinely make decisions about content removal, account suspension and the visibility of information. Such decisions have significant consequences for users. These decisions can bring in the restriction of speech, reputational damage or exposure to harassment and harmful content. Users need to have a mechanism to seek remedial actions from the digital platforms.

The UNESCO Guidelines for the Governance of Digital Platforms (2023) emphasize the importance of transparent, accessible and effective user complaint systems. These systems should allow individual users to report harmful content, appeal moderation decisions and seek redress when platform policies are applied unfairly or inconsistently. Grievance mechanisms should also ensure fairness, timeliness and respect for human rights.

Independent oversight is another important component of platform accountability. Oversight mechanisms may include regulatory bodies, judicial review processes, independent advisory boards or multistakeholder oversight institutions which help ensure that platform decisions are subject to scrutiny and that users' rights are protected.

Civil society organizations play an important role in strengthening grievance and oversight systems. By documenting user experiences, monitoring complaint processes and advocating for stronger oversight mechanisms as well as supporting the independent oversight, civil society can help ensure that digital platform governance remains transparent, accountable and responsive to users' concerns.

This unit is suitable for legal professionals, digital rights advocates, regulatory authorities and technical developers or engineers working at the intersection of technology and rights.

It centers on enhancing complaint and grievance mechanisms while promoting a human rights by design approach in the development and deployment of digital platforms and technologies (OHCHR, 2011; UNESCO, 2026)

What this unit provides

- *Technical Guidance: Offers a 5-step framework for "Human Rights and Safety by Design" for engineers and data scientists to mitigate risks before deployment.*
- *Grievance Standards: Defines the core principles of an effective complaint mechanism, focusing on accessibility in local languages, transparency, and the right to appeal.*
- *Oversight Advocacy: Supports legal professionals in advocating for independent regulatory authorities that are not under direct state control.*

What can civil society do?

Civil society organizations can support effective grievance mechanisms by helping users understand their rights and platform reporting systems.

- Guide users on how to report harmful content and challenge moderation decisions.
- Help users access available remedies and complaint processes.
- Monitor whether platforms respond promptly to complaints and appeals.
- Assess whether appeals systems are transparent, and outcomes clearly communicated.
- Identify gaps in user protection and platform accountability.
- Advocate for independent regulatory authorities and stronger oversight frameworks.
- Promote judicial review of platform decisions and multistakeholder oversight involving civil society and academia.
- Collect data on complaints, appeals, and cases.
- Analyze patterns of governance failures to support advocacy, research, and policy recommendations.

BOX 1: HUMAN RIGHTS AND SAFETY BY DESIGN IN TECHNICAL DEVELOPMENT

Digital platforms rely on complex technological systems such as algorithms, artificial intelligence and automated moderation tools. The design and deployment of these systems can have significant implications for freedom of expression, privacy, equality and user safety (UNICEF, 2021)

To promote innovation, engineers, data scientists and technical professionals should integrate human rights considerations and safety-by-design principles throughout the lifecycle of digital products and services. This means identifying risks early, designing systems that minimize harm and continuously monitoring their real-world impacts.

Steps for Promoting Human Rights in Technical Work

Step 1: Risk Assessment

- Identify potential human rights impacts at each stage of product or service development.
- Consider vulnerable and marginalized groups in assessing the effects of features, algorithms and data use.
- Advocate for Watermarking Obligations and Provenance Tools to ensure that all AI-generated content is clearly identifiable and its origin can be traced.

Step 2: Ethical and Human Rights-Informed Design

- Integrate privacy, security, fairness and inclusivity into system design.
- Apply principles of ethical AI and ethical technology development.

Step 3: Testing and Simulation

- Conduct scenario testing to identify unintended consequences including mandatory pre-launch risk assessments for all new generative AI features.
- Evaluate algorithms for bias, fairness and discriminatory outcomes before deployment.

Step 4: Continuous Monitoring

- Monitor platform performance and user interactions to detect new risks or harm.
- Implement automated and manual checks to ensure compliance with ethical and human rights standards.

Step 5: Collaboration and Reporting

- Work with civil society, human rights experts and regulatory bodies to review platform practices.
- Maintain transparent reporting mechanisms for findings, risks and mitigation measures.
- Adjust systems and policies based on insights from external audits and feedback.

BOX 2: EFFECTIVE USER GRIEVANCE MECHANISM

An effective user grievance mechanism allows individuals to report harmful content, challenge platform decisions and seek remedies when their rights are affected. For this mechanism to be fair and effective, they should be designed according to principles of accessibility, transparency and accountability.

Accessibility

Users should be able to easily submit complaints through clear and user-friendly reporting channels. Reporting tools should be available in local languages and accessible to people with different levels of digital literacy.

Transparency

Platforms should clearly explain how complaints are reviewed, what criteria are used for decisions and how long the process may take. Users should be informed about the status and outcome of their complaints.

Timeliness

Grievance mechanisms should respond to user complaints promptly within a reasonable timeframe.

Fair and consistent review

Moderation and appeal decisions should be applied consistently across different users, regions and languages.

Right to appeal

Users should have the opportunity to challenge platform decisions such as content removal, account suspension or other moderation actions.

User protection

Grievance mechanisms should prioritize the safety of victims of harassment, threats or harmful content while protecting privacy and personal data.

Independent oversight

Platforms should allow independent oversight such as external review bodies, regulatory authorities or multistakeholder mechanisms to assess whether grievance mechanisms are fair.

Action Points for Civil Society

Civil society organizations, media institutions and digital rights groups can take the following actions to strengthen user grievance mechanisms and oversight of digital platforms.

1. Supporting users

a) Provide guidance to users on how to report harmful content or abusive behaviour to digital platforms.
b) Assist individuals in using platform appeals processes when content is removed or accounts are suspended.
c) Help users understand their digital rights and available remedies when platform decisions affect them.

2. Monitoring grievance mechanisms

a) Examine whether digital platforms provide accessible and transparent complaint systems.
b) Monitor how quickly platforms respond to user reports and appeals.
c) Identify patterns of unresolved complaints or inconsistent enforcement of platform rules.
d) Study the effectiveness of the grievance mechanisms.

3. Documentation

a) Collect data on user complaints, appeals and outcomes of moderation decisions.
b) Collect lived experiences of individual users on complaints handling.
c) Document cases where grievance mechanisms fail to address harmful content or unfair moderation practices.
d) Publish reports with issues in platform grievance processes.

4. Advocating for independent oversight

a) Advocate for independent oversight bodies that can review platform governance practices.
b) Promote the development of transparent regulatory frameworks that protect user rights while respecting freedom of expression.

c) Encourage governments to adopt multistakeholder approaches to digital platform oversight.

5. Engagement with digital platforms

a) Engage with platform representatives to communicate concerns about grievance mechanisms and moderation practices.
b) Advocate for clearer reporting systems, stronger appeals processes and improved communication with users.
c) Encourage platforms to publish data on complaint handling and moderation outcomes.

6. Public Awareness

a) Inform the public about their rights in digital spaces and available complaint mechanisms.
b) Publish explanatory guides that help users understand how platform moderation systems work.
c) Promote public discussion about the importance of transparency and accountability in digital platform governance.
d) Enforce mandatory pre-launch human rights risk assessments for all new Generative AI features. Enforce/advocate for mandatory "Human Rights Risk Assessments" and "Age-Appropriateness Testing" before new AI features or algorithms are launched in the regional market.

Implementation Checklists, Tools and Monitoring Indicators for User Grievance & Oversight

Implementation Checklist	Implementation Tools	Monitoring Indicators
1. Grievance System Accessibility & Transparency: Test how easily users can report harm and evaluate if responses are meaningful (human) or generic (automated bots).	UX Accessibility Scorecard: A rating tool to evaluate: (1) Ease of finding the report button, (2) Local language availability, and (3) Clarity of human vs. bot responses.	Average ease-of-use score given by users % of responses that are personalized vs. generic bot replies.

Implementation Checklist	Implementation Tools	Monitoring Indicators
2. Support for User Appeals & Redress: Assist users who have been wrongly banned or silenced in navigating the platform's complicated appeal and contestation processes.	Appeals Template Letter: A professional form letter citing specific platform community standards and international rights to contest wrongful account suspensions.	% of wrongful bans overturned or content restored after civil society intervention.
3. Safety-by-Design & Technical Accountability: Audit local apps and train technical professionals to ensure human rights and safety are built into the code before a product is launched.	Safety-by-Design Matrix: A 5-step risk assessment tool covering: Risk Assessment, Ethical Design, Scenario Testing, Continuous Monitoring and Collaboration.	# of technical audits shared with developers # of engineers/ developers trained in Safety-by-Design principles.
4. Response Timeliness for High-Risk Harm: Monitor how quickly platforms act on life-threatening reports.	Critical Response Clock: A tracking method to measure the exact time elapsed from the moment a report is filed to the platform's final action.	Average response time taken by platforms to make decision.
5. Independent Regulatory & Oversight Advocacy: Campaign for oversight bodies that are free from state control and ensure they include diverse representatives from civil society and academia.	Regulatory Autonomy Scorecard: A checklist to measure if the national regulator is free from political interference and follows the UNESCO 2023 Principles.	Existence of an independent oversight body (Yes/No) # of CSOs/Academic experts included on advisory boards.
6. Documentation of Failures: Document and publish instances where platform grievance mechanisms failed to protect users or ignored high-risk reports.	Grievance Failure Log: A data-base recording cases where reporting tools were broken, languages were unsupported or appeals were ignored.	# of failure case studies published by CSOs to pressure platforms for systemic improvements.

Unit VI

Conclusion: Strengthening Civil Society Engagement in Digital Platform Governance

Digital platforms are the central infrastructures of contemporary communication determining how information circulates, how public debate happens and how individuals exercise their rights to freedom of expression and access to information. However, the growing influence of these platforms has brought in complex governance challenges.

Across Bangladesh, Nepal, and Sri Lanka, these challenges are emerging within rapidly evolving regulatory environments. Governments are increasingly attempting to regulate digital platforms, often in response to concerns about misinformation, online harms and threats to democratic processes. However, regulatory approaches remain fragmented and, in some cases, undermining freedom of expression or concentrating excessive state control over digital communication. Global technology companies continue to exercise significant influence over digital information ecosystems through opaque algorithms, content moderation systems and advertising models (Digital Rights Nepal, n.d.). They are often not present or unresponsive in Global Majority countries with limited markets in the matters of digital governance. To sustain this work, civil society should advocate for independent funding models, such as fractional platform levies on local digital ad revenues, to support independent oversight, algorithmic audits, and multilingual moderation without relying on government budgets.

Effective digital governance depends on the active engagement of diverse stakeholders, including civil society, media, academia, technical professionals, educators and regulators, in the development, implementation and operation of governance systems. Ensuring these systems are inclusive, accountable and responsive requires fostering both collaboration and dialogue.

Collaboration and dialogue should be actively encouraged through structured, constructive discussions, deliberations and knowledge sharing, allowing stakeholders to exchange ideas, expertise and experiences. Establishing working groups, task forces or advisory committees provides meaningful opportunities for participation, enabling stakeholders to contribute to shaping regulatory proposals, platform policies

and ethical practices. Such inclusive and cooperative approaches ensure governance frameworks are transparent, accountable, aligned with human rights principles and responsive to the needs of all users, fostering a resilient and sustainable digital ecosystem.

In this context, civil society organizations play a vital role in promoting accountable and rights-respecting platform governance. Their work in monitoring online harm, verifying information, documenting abusive behaviour, supporting victims of digital abuse and advocating for transparent platform policies helps strengthen democratic resilience in digital environments.

The UNESCO Guidelines for the Governance of Digital Platforms (2023) provide an important framework for addressing these challenges. The guidelines emphasize a human-rights-based, multistakeholder approach that balances the need to address harmful content with the protection of fundamental freedoms. Transparency, accountability, safety by design, and inclusive governance are central principles that should guide both regulatory frameworks and platform practices.

There are several practical areas where civil society engagement is important. These include protecting freedom of expression and access to information, identifying harmful online behaviour and its root causes, strengthening fact-checking and watchdog journalism, promoting transparency and accountability of digital platforms and ensuring accessible grievance mechanisms for users. Together, these actions contribute to more platform governance and healthier digital information ecosystems.

Moving forward, effective platform governance will require collaboration among governments, digital platforms, civil society organizations, media institutions, researchers and international stakeholders. It emphasizes joint planning, shared responsibilities, and coordinated action in designing, implementing, and refining governance frameworks and platform policies. For countries in Global Majority with limited resources, expertise and technology, it is worthwhile to engage with digital platforms to develop mutually beneficial governance modals in line with principles of UNESCO guidelines. Civil society actors can play a bridging role by facilitating dialogue between these stakeholders, generating evidence-based insights and advocating for policies that uphold democratic values and human rights. To ensure the independence of these governance systems, CSOs should advocate for sustainable funding models, such as Platform Levies (a small percentage of local digital advertising revenue), to fund regulator training and independent audits without relying on shifting government budgets.

At the same time, digital platforms themselves must integrate ethical design and safety-by-design principles into their technological systems, and do their optimum

efforts for algorithmic transparency, meaningful user protection and human rights respecting data practices. These are essential to ensuring that digital technologies serve the public interest rather than undermine it. They also need to proactively seek collaboration with state institutions in Global Majority to extend support on expertise and technology to minimize harm on the public.

The goal of digital platform governance should not only be to control harmful content but also to build trustworthy, inclusive and rights-respecting digital environments.

Unit VII

Way Forward: Practical Guide to Integrating the Toolkit for Civil Society

This section, a practical guide, to integrating the Toolkit for Civil Society is developed in the context of South Asia based on the **UNESCO Guidelines 2023** into university curricula and other institutions. It aims at contextualizing the guidelines in Bangladesh, Nepal, and Sri Lanka, where resources are limited, political sensitivities exist, and digital challenges such as MDM, linguistic diversity, elections information integrity are acute.

Why the Toolkit Fits Education

The toolkit fits to educational purpose. Because has clear Units (I–VII), with practical scenarios checklists, tools and monitoring indicators for harmful contents, and a strong emphasis on **human rights due diligence**, **multistakeholder approaches**, **transparency**, and **systemic oversight** rather than heavy-handed content policing. It aligns with UNESCO's own work on **Media and Information Literacy (MIL)**, **AI competency frameworks** for students/teachers, and ethical AI integration in higher education. It can help build "process auditors", future regulators, journalists, lawyers, tech professionals, and informed citizens, who understand platform accountability without becoming censors (Acharya & Shrestha, 2026-a).

Step-by-Step Integration into University Curricula

1. Embed as a Standalone or Core Module

- **Create a 3-credit elective or required course titled something like "Digital Platform Governance and Civil Society Engagement" or "Regulating the Digital Ecosystem for Bangladesh, Nepal and Sri Lanka."**
- **Target audience:** Undergraduate, graduate and post-graduate students in Media Studies and other branches of Humanities and Social Sciences, Law and Human Rights Studies, Computer Science/IT and Public Policy.

- **Structure around the toolkit's Units**
  - **Unit I (Supporting Freedom of Expression and Access to Information) + Annexes → Core reading on civil society's role to implement checklists tools and monitoring indicators.**
  - **Units II–V (Addressing harmful content online, Identification of Abusive behaviours and root causes, Fact-checking and watchdogging digital platforms, Promoting transparency and accountability of digital platforms) → Modules on transparency reporting, algorithmic audits, and harmful content classification.**
  - **Unit VI (User grievance mechanism and independent oversight) + action points for civil society.**
  - **Unit VII (Conclusion:** Strengthening Civil Society Engagement in Digital Platforms Governance) + Case studies from Bangladesh, Nepal and Sri Lanka + Key terminologies.
- **Teaching methods:** Case studies from Bangladesh (BTRC challenges), Nepal (Press Council critiques), Sri Lanka (Online Safety Act delays); role-playing regulator-platform-civil society negotiations; group audits using the toolkit's checklists. Furthermore, fact-checking items, digital governance report, and AI related studies from Nepali CSOs.

2. Transdisciplinary Approach

- **Law/Policy programs:** Use as core texts for courses on cyber law, civil society engagement, media regulation, or data protection — pair with ICCPR three-part test and global comparisons such as EU AI Act, UNESCO AI Ethics.
- **Media Studies and other branches of Humanities and Social Sciences:** Link to MIL — teach students to analyze platform moderation failures, fact-check deepfakes, and advocate for user empowerment.
- **Computer Science/IT:** Add technical modules on algorithmic accountability, watermarking/C2PA standards, red-teaming AI moderation systems, and human oversight of automated tools.
- **Education/Faculty Development:** Train future teachers using UNESCO's AI Competency Frameworks alongside the toolkit for classroom discussions on ethical AI use and combating disinformation.
- **Cross-cutting:** Make it part of general "Digital Citizenship" or "Ethics in the Digital Age" requirements for all levels of university education.

3. Practical Learning Activities

- **Simulate harmful online content using Box 1, Box 2, Implementation Checklists, Implementation Tools and Monitoring Indicators integrating the Action Points for civil society.**
- **Student projects:** Conduct exercise to analyze the contents using Potters' 7 skills, Identification of misinformation, disinformation and malinformation following the Fact-checking Steps (Acharya & Shrestha, 2026-b).
- **Guest lectures from CSOs working in the field of research, monitoring, campaign and advocacy in Bangladesh, Nepal and Sri Lanka, followed by toolkit-based debriefs.**
- **Capstone:** Multistakeholder Policy Lab where students propose adaptations of the toolkit for their country's context.

4. Faculty and Institutional Support

- **Offer short faculty workshops for 1–2 days on using the toolkit, co-organized with UNESCO's regional and national office or local CSOs including digital rights groups.**
- **Update existing syllabi gradually, start as optional readings, then adopt as mandatory.**
- **Partner with UNESCO's South Asia initiatives, e.g., AI Readiness Assessments in Bangladesh, digital learning policies in Nepal, for resources and credibility. (Freedom Forum, 2025-b)**

Integration Beyond Universities

CSOs and Training Institutions and Academia

- **Mandatory interaction/refresher training for CSOs, in Bangladesh, Nepal and Sri Lanka.**
- **Use Units I–III for supporting freedom of expression and access to Information, addressing harmful content online, Identification of abusive behaviours and root causes, Unit IV for fact-checking and watchdogging digital platforms, Unit V for the promotion of transparency and accountability of digital platforms, and Unit VI for the user grievances mechanisms and independent oversight.**

CSOs including Rights Advocacy Groups

- **Capacity-building workshops for digital rights organizations, fact-checkers, and journalists, focus on using checklists to monitor platforms and advocate for better policies.**
- **Integrate into MIL programs for communities, youth, and vulnerable groups such as women, ethnic minorities, empower them as watchdogs per UNESCO guidance.**

Media Training Institutes

- **Embed in professional development courses:** Teach reporters how platforms moderate content and how to push for transparency without self-censorship.

Libraries, Community Centers & Lifelong Learning

- **Short public sessions or MIL clubs using simplified toolkit materials, e.g., harmful content table translated into Nepali, Bangla, Sinhala.**
- **Libraries as neutral hubs for digital ecosystem reports discussions.**

Realistic Implementation Roadmap for Bangladesh, Nepal and Sri Lanka

- **Phase 1 (Short-term, 6–12 months):** Pilot in 2–3 universities, e.g., Tribhuvan University in Nepal, Jahangirnagar University in Bangladesh and University of Vocational Technology in Sri Lanka. Run one faculty workshop and one student simulation based on the translated version of the toolkit in local language.
- **Phase 2 (1–2 years):** Scale via national education ministries or University Grants Commission equivalents. Link to existing digital transformation efforts in higher education. Explore support from UNESCO, international donors, or platform.
- **Phase 3 (Ongoing):** Annual reviews, update with new AI developments such as watermarking, red-teaming. Measure success via student/regulator feedback, not just exam scores: Are participants better at spotting systemic risks?

Potential Challenges and Coping Strategies

- **Resource constraints:** Start low-cost, use the open toolkit PDF and free UNESCO

MIL resources. Blend online and offline module.

- **Political sensitivities:** Frame as “capacity building for digital governance” aligned with national laws and human rights, not anti-government critique. Emphasize multistakeholder balance.
- **Relevance:** Supply sufficient local examples, e.g., election disinformation in Nepal, linguistic moderation challenges in Bangladesh.
- **Evaluation:** Use pre/post knowledge tests on the five UNESCO principles, plus qualitative feedback on practical skills.

Annex 1

Case Study 1. Bangladesh

The spread of lies and harmful content on digital platforms is a common phenomenon in the Bangladeshi digital ecosystem, as it is elsewhere in the world. The results of the dissemination of disinformation and fake news on digital platforms, including social media, are often disturbing, and they affect human lives and property. The following are two cases in which journalism organizations, religious places, and people were affected by the spread of lies on digital platforms, particularly social media. These two incidents drew national and international attention.

Case 1: Attack on Prothom Alo and The Daily Star

The offices of the country's two leading media organizations – *Prothom Alo and The Daily Star came under violent mob attacks on 18 December 2025. An investigation jointly conducted by The Daily Star and Dismislab, a fact-checking organization, reveals that threats against both media outlets had been spreading on social media for months. Days before the attacks were made, these two media outlets were being labelled as "Indian agents" and "anti-national forces" on social media by many social media users. They even made calls to "burn and attack" the offices of these two media outlets (Amnesty International, 2026). According to a NewAge news report, a mob entered the Prothom Alo office at around 11:00pm and vandalized it. Later, around 1:15am, the mob moved to the Daily Star office in the capital and they vandalized glass panels, dragged out chairs, tables, and papers and set them on fire (NewAge, 2025).*

Case 2: Attacks on Buddhist monasteries, shrines and Buddhist houses

On September 29, 2012, an image depicting the desecration of a Quran, was posted on the wall of a Buddhist person's Facebook profile. It triggered a series of attacks on Buddhist monasteries, shrines and Buddhist houses in Cox's Bazar, Bangladesh. Investigation later found that the picture that instigated the violence against the Buddhist community had actually been doctored (Business Standard, 2019). The BBC reports that an agitating mob set fire to homes and temples of the Buddhist community, forcing families to flee. A curfew was imposed as part of security measures. Police said the man accused of posting the image was tagged in the photo but did not post it himself on Facebook (BBC, 2012; Business Standard, 2019).

2. Nepal

The COVID-19 pandemic, between 2020 to 2023, was not only a public health, but it was also an information integrity crisis as disinformation on digital platforms showed real-world impacts for vulnerable communities. The utilization of unverified herbal remedies and alternative treatments during the COVID-19 period illustrates the repercussions of false information in Nepal (Lamichhane, 2024). But more concerning was how disinformation targeted migrant workers and Muslim communities.

At the time, according to the 2021 census, 52% of the population were internet users, with Facebook being the dominant platform as 79.8% of respondents who encountered misinformation reported seeing it on Facebook (Acharya, 2023). The high number of users with limited media and information literacy and awareness about misinformation added fuels to fear, confusion and harming vulnerable.

CASE 1: Returnee migrants targeted online and offline

Nepal imposed lockdown in March 2020 following other countries around the world to stop spread of coronavirus. Many Nepali labour migrants began returning home from India, the Gulf and elsewhere. But they were met with fear, suspicion and online hostility that turned into real-world discrimination and exclusion.

A 19-year-old Nepali student who returned to Nepal from France tested positive for the virus, and the people on social media accused her of spreading the virus and openly posted death threats against her (Kathmandu Post, 2020). The Kathmandu Post (2020) citing police officials reported that around three online harassment cases related to coronavirus were being filed daily and people who had returned from abroad were being called 'corona patients' online and receiving harsh comments from anonymous users.

The disinformation around returnees was fueled by various actors. Misinformation was fueled by social media platforms, especially Facebook and YouTube, during the initial months of the lockdown, creating a moral panic which led to returnee migrants and Muslim populations experiencing fear and social isolation (Regmi et al., 2022). Political actors added to the problem, for example Prime Minister K.P. Oli repeatedly used the term 'unchecked Nepali migrants' as responsible for increased infections and at the local level, authorities marked houses returnee migrants to enforce quarantine (Dhungana, 2020).

The situation women returnees were worse. Media and social platforms were used to make degrading and stigmatizing comments toward returning pregnant migrant women, hurting their self-respect and contributing to spreading negativity among the general public toward them (National Human Rights Commission of Nepal, 2020).

CASE 2: Muslim communities blamed as spreaders of coronavirus

The other wave of COVID-related disinformation in Nepal targeted Muslim minority communities. Stigmatizing behaviour against Muslims and returnee migrants spiked during the initial phase of COVID-19, partly due to the early COVID-19 positive cases found in these population groups (Regmi et al., 2022).

An example of two Muslim women who unintentionally dropped banknotes while walking on the street were projected as dropping spitted banknotes to spread the virus. They were subjected of arrest, quarantine and harassment before being freed after tests were negative. Anti-Muslim rhetorics was exacerbated by cross-border disinformation. Some Indian television channels spread false and unverified claims falsely accusing a Nepali Muslim man of plotting to spread the virus in India (Acharya & Shrestha, 2026). Despite official denials, the channels repeated the misinformation, amplifying conspiracy narratives and communal suspicion across the open Nepal-India border (Acharya & Shrestha, 2026).

Research on the lived experience of Nepal's Muslim community during this period documented how media and social media content enforced a moral panic, promoting a negative public perception of Muslims as potential virus carriers and spreaders, and how this had real implications for their wellbeing and relationships within communities, including perceived fear, othering practices such as stigma, discrimination and humiliation (Regmi et al., 2022).

How did the platforms fail? A local language moderation gap

There was the absence of effective content moderation in the Nepali language on digital platforms in both cases. In Nepal, where Nepali is the primary language, there is a lack of language experts for platform content moderation; without such expertise, the nuances of language are ignored and what is quite obviously harmful content has been allowed to flourish (Adhikari, 2024). This gap is not incidental; it shows the pattern in which digital platforms under-invest in moderation infrastructure for languages spoken by smaller or lower-income populations.

The government's response to platform failures was also not enough. There was no effective mechanism by which civil society or government bodies could formally flag harmful Nepali-language content to platforms and receive guaranteed action.

How did civil society respond? NepalFactCheck.org

In the absence of platform action, civil society stepped in. NepalFactCheck.org was initiated in March 2020 as an urgent response to COVID-19 misinformation, a joint initiative of the Center for Media Research – Nepal (CMR-Nepal) and MySansar, Nepal's first Nepali-language blog. The initiative was launched because pseudoscience and false claims were being circulated by opinion leaders and influencers, and big platforms were noncommittal in combating misinformation, while the media system was weakened by the struggle to survive (Acharya & Shrestha, 2026).

In its first five years, NepalFactCheck.org published 381 fact-checks, with the early COVID period constituting one of three peak periods of activity (CMR-Nepal, 2025). Crucially, the organization operated in Nepali language. However, the initiative faced significant structural limitations: lack of funding, limited staffing and the absence of a formal channel to escalate flagged content to platforms. This stands in contrast to the "bridging dialogue" mechanism described in the Sri Lanka case study, where a CSO's structured engagement with Meta's regional teams resulted in content removal within 24 hours.

Lessons learned

1. Disinformation caused measurable real-world harm.
2. The Nepali-language moderation gap is evident during the spread of harmful content. The platforms lacked the language capacity and the accountability mechanisms to respond to locally embedded harm.
3. Civil society can fill a vacuum, at least by some measure, by initiating actions without enough resources and digital platform collaboration.
4. Political actors can amplify disinformation environments.
5. Cross-border disinformation requires cross-border platform accountability.

Nepal case shows the urgent need for CSOs to build engagement with platforms, and for platforms to make commitments to invest in local language moderation capacity especially during crisis periods.

3. Sri Lanka

Case 1: Freedom of Expression Under Strain in Contemporary Sri Lanka

Freedom of expression in Sri Lanka has come under increasing scrutiny in recent years, particularly in the context of political dissent, artistic expression, and digital communication. This case study examines four key instances: the media restrictions during the 2022 anti-government protests, the detention of journalists and activists for online speech, and two emblematic cases involving comedian Nathasha Edirisooriya and writer Shakthika Sathkumara. Together, these cases illustrate a broader pattern in which existing legal frameworks—especially those ostensibly designed to protect public order or prevent hate speech—are used in ways that risk undermining constitutionally guaranteed freedoms.

Media Blackout During the 2022 Anti-Government Protests

The 2022 "Aragalaya" protests represented a historic moment of mass civic mobilization in Sri Lanka, with demonstrators demanding political accountability amid economic collapse. However, the state's response raised serious concerns about freedom of expression. Authorities imposed emergency regulations and restricted media access, effectively creating a partial media blackout during critical moments of the protests.

In July 2022, protesters occupying Galle Face Green were forcibly evicted under emergency regulations. This action was later challenged before the Supreme Court. In a landmark 2025 judgment, the Court held by a 2–1 majority that the use of emergency powers to disperse a nonviolent protest violated fundamental rights, including freedom of expression and peaceful assembly (Balasuriya, 2025; Supreme Court of Sri Lanka, 2022).

This ruling affirmed that political protest constitutes protected expression under the Constitution. However, the initial suppression combined with limitations on media reporting demonstrates how executive powers can be used to curtail information flows during politically sensitive periods.

In early 2025, U.S. President Donald Trump claimed on his social media platform that USAID (the United States Agency for International Development) had been used to "spread chaos" and "change governments," specifically naming Sri Lanka as a recipient of such funds. Trump's post alleged that "George Soros received $260,000,000.00 from USAID and used this money to spread chaos in Sri Lanka". This claim has been

weaponized by local political actors in Sri Lanka to retroactively frame the 2022 protests as a foreign-orchestrated color-revolution, providing a political justification for the ongoing surveillance of civil society and independent media.

Detention of Journalists and Activists for Online Expression

Alongside restrictions on traditional media, authorities also targeted individuals for their online speech during and after the protests. Journalists, activists, and ordinary citizens faced arrest or intimidation for social media posts perceived as critical of the government.

These actions reflect a broader trend of digital surveillance and criminalization of dissent, where online platforms—once seen as spaces for democratic participation—become sites of legal vulnerability. The use of laws such as the International Covenant on Civil and Political Rights (ICCPR) Act and provisions of the Penal Code have enabled authorities to frame dissenting speech as threats to public order or national harmony. This has had a chilling effect, discouraged open political discussion and weakened democratic accountability.

The Case of Nathasha Edirisooriya

The arrest of stand-up comedian Nathasha Edirisooriya in May 2023 highlights the precarious position of artistic expression in Sri Lanka. In her show *Modibhimanaya ("Fool's Pride"), she made satirical remarks about Buddhist beliefs. Although she removed the content and issued an apology after public backlash, she was arrested and detained for several weeks under the ICCPR Act on charges related to incitement of hatred (Global Freedom of Expression, 2025).*

In July 2023, the Colombo High Court granted her bail, citing insufficient grounds for continued detention (Wickramasinghe, 2023). The Human Rights Commission of Sri Lanka later found that her arrest and travel restrictions violated constitutional protections, including freedom of speech and equality before the law (Global Freedom of Expression, 2025). This case underscores the expanding use of the ICCPR Act beyond its original purpose of preventing hate speech, raising concerns that it is being applied to suppress controversial but legitimate forms of expression.

The Case of Shakthika Sathkumara

The case of Shakthika Sathkumara further illustrates the risks faced by creative writers. In April 2019, Sathkumara was arrested for a short story titled *Ardha ("Half"),*

which he had shared on Facebook. The story explored sensitive themes, including homosexuality and abuse within a Buddhist monastic setting. Following complaints by religious groups, he was detained for 127 days under the ICCPR Act and the Penal Code (Amnesty International, 2021).

Although the Attorney General eventually dropped the case in 2021 without indictment (PEN International, 2025), the United Nations Working Group on Arbitrary Detention found that his arrest violated international law and called for compensation (Spector, 2020). This case is particularly significant because it involves purely literary expression shared online, demonstrating how artistic content can be reclassified as criminal speech under broad legal interpretations.

Conclusion

These cases collectively reveal a concerning trajectory in Sri Lanka's approach to freedom of expression. While the Constitution formally guarantees this right, its practical realization is increasingly constrained by the expansive use of existing laws and regulations. The ICCPR Act, emergency regulations, and provisions of the Penal Code are being deployed not only to address genuine threats such as hate speech or violence, but also to target dissent, satire, and artistic expression.

The result is a chilling effect on public discourse, where individuals may self-censor to avoid legal repercussions. The Supreme Court's recognition of protest as protected expression offers an important counterbalance, yet systemic challenges remain. Overall, there is a clear and growing tendency toward using established legal frameworks as tools to curtail, rather than protect, freedom of expression in Sri Lanka.

Case 2: Bridging the Context Gap – Monitoring Coordinated Hate Speech

Context:

This case is situated in Sri Lanka's recent election cycles (particularly around 2019–2020), a period marked by heightened political polarization, ethnic sensitivities, and increased online activity. In such moments of national tension, social media platforms, especially Facebook, became key spaces for political messaging and public discourse.

However, platform moderation systems struggled to effectively interpret Sinhala and Tamil language content, particularly when harmful narratives were embedded in coded language, historical references, and culturally specific expressions. As a result, hate speech and divisive content were able to circulate widely and rapidly before being detected by automated AI systems. This created a significant "context gap", where harmful intent was visible to local communities but not to platform algorithms or moderators operating without local knowledge.

Monitoring (The Task)

In response to this challenge, a local Civil Society Organization (CSO) adopted a Narrative Audit Tool to systematically monitor online discourse. Moving beyond simple keyword searches, the organization identified 15 recurring hate tropes, patterns of coded language, historical myths, and dehumanizing metaphors used in Sinhala and Tamil to subtly incite hostility against a specific ethnic minority group. This approach allowed the CSO to capture not just explicit hate speech, but also implicit and context-driven forms of harm that often evade automated detection systems.

Documentation

The monitoring team collected approximately 200 time-stamped screenshots as evidence, ensuring that each instance of harmful content was properly recorded and verifiable. By analyzing metadata and tracking how posts were shared and amplified, they uncovered a clear pattern: nearly 60% of the most viral harmful narratives originated from a small, coordinated network led by three prominent influencer accounts.

This shifted the understanding of the issue, from scattered incidents to a structured and coordinated campaign, supported by identifiable actors.

Engagement

Recognizing that evidence alone is not enough, the CSO engaged directly with Meta's Regional Public Policy and Safety teams through an established "bridging dialogue" channel. Instead of submitting generic reports, they prepared a confidential evidence memo that translated the identified hate tropes into clear explanations of their cultural meaning and real-world implications. This step was crucial in helping platform moderators—who may lack local linguistic and cultural expertise—understand how the content violated Violence and Incitement policies.

Outcome (Effectiveness)

With structured evidence and contextual clarity, the platform conducted a manual review of the identified accounts and content. Within 24 hours, the three key influencer accounts and their coordinated network of harmful posts were removed. This rapid response significantly reduced the spread of harmful narratives and helped prevent the escalation of online hate into offline tensions or violence.

Key Lesson for Practitioners

This case highlights that context matters as much as content. Automated systems alone are not sufficient in multilingual and culturally complex environments.

- **Monitoring helps identify patterns of harm**
- **Documentation builds credible evidence**
- **Engagement ensures action**

The real impact lies in bridging the gap between local knowledge and global platform systems.

Source Note:
Adapted from the *Social Media Monitoring of the 2019 Presidential Election* report by Hashtag Generation in partnership with Democracy Reporting International (DRI) and the Centre for Policy Alternatives (CPA). Full methodology and reports are available at *HashtagGeneration.org* and *CPALanka.org*.

Annex 2

Good Practices of Civil Society for Digital Platform Governance

This annex section presents selected best practices of civil society engagement in digital platform governance from America, Europe, Asia, and Africa, as outlined below:

1. Brazil's Marco Civil da Internet (The Internet Bill of Rights)

This is widely considered the gold standard for civil society-led policy development. Rather than the government drafting laws in isolation, the Center for Technology and Society of the Getulio Vargas Foundation (FGV) led a multistakeholder process. They used an online platform to crowdsource contributions from citizens, academics, and NGOs to draft a human-rights-respecting framework for platform liability and net neutrality (Center for Technology and Society, 2014).

Role of Civil Society: Co-drafting legislation and facilitating public consultation.

2. USA: Ranking Digital Rights (RDR)

RDR is the best practice for "Watchdogging and Transparency Auditing" They produce an annual *Corporate Accountability Index* that evaluates the world's most powerful digital platforms on their disclosed policies regarding freedom of expression and privacy. Their methodology forces platforms to be more transparent about how they moderate content and handle user data (Ranking Digital Rights, 2022).

Role of Civil Society: Independent auditing and benchmarking of platform transparency.

3. South Africa's Real411 (Media Monitoring Africa)

Real411 is the best practice for User Grievance Mechanisms and Addressing Harmful Content. Developed by the NGO Media Monitoring Africa in partnership with the Electoral Commission, it is a platform where citizens can report digital harms such as misinformation and hate speech. A panel of independent legal experts reviews the reports and communicates directly with social media platforms to facilitate takedowns or labels (Media Monitoring Africa, 2021).

Role of Civil Society: Serving as an independent oversight body and trusted flagger.

4. European Digital Rights (EDRi) and the DSA Coalition

During the drafting of the European Union's **Digital Services Act (DSA)**, EDRi (a network of over 45 NGOs) led the People vs Big Tech coalition. They conducted Legislative Audits and lobbied the European Parliament to ensure the law included protections for anonymous speech and limited the power of 'Recommender Systems' (algorithms) to amplify harmful content (European Digital Rights, 2022).

Role of Civil Society: Policy advocacy and human rights impact assessments of regional law.

5. Investigative Watchdogging and Disinformation Mapping (Philippines' Rappler)

In the Philippines, the CSO and media organization Rappler (founded by Nobel Laureate Maria Ressa) pioneered the use of data science to hold digital platforms accountable for patriotic trolling and state-sponsored disinformation. By analyzing millions of social media data points, they mapped the weaponization of the internet and provided evidence to platforms like Facebook (Meta) to shut down coordinated inauthentic behaviour (CIB) networks. This research shifted the conversation from fake news to networked disinformation, forcing platforms to update their community standards in the region (Ressa, 2022).

Role of civil society: Big data analysis and network mapping using data to identify threat actors and hold platforms accountable for systemic harms.

Annex 3

Key Terminologies

Ad library: A publicly accessible database provided by digital platforms where users can track advertisements on the platform, including sponsors and target audiences (Meta, n.d.; European Union, 2022).

Algorithmic amplification: The way digital platform algorithms prioritize certain content to maximize user engagement (UNESCO, 2023).

Black box: A term describing the opaque nature of platform algorithms where the input and output are visible, but the internal decision-making logic remains secret (Pasquale, 2015).

Content moderation: The process, whether manual or automated, through which platforms monitor, review, remove, label, or de-prioritize content according to community standards and legal requirements (Gillespie, 2018).

Coordinated Inauthentic Behaviour: Multiple accounts, often bots or fake profiles, working together to amplify content in order to mislead people or manipulate public discourse for political or commercial goals (Meta, n.d.).

Disinformation: False or misleading information deliberately created and disseminated with the intention to deceive, manipulate, or cause harm (Wardle & Derakhshan, 2017).

Fact-checking: The systematic verification of publicly available statements or information to determine whether they are accurate or false (International Fact-Checking Network, n.d.).

Global majority: A term referring to the majority of the world's population living outside the Global North or highly industrialized countries (Campbell-Stephens, 2020).

Hate speech: Content that attacks, disparages, or incites hostility against a group based on identity characteristics such as caste, ethnicity, gender, religion, or race (United Nations, 2019).

Human Rights Due Diligence: A proactive process to identify, prevent, mitigate, and account for how platform activities may negatively affect human rights (United Nations, 2011).

Human Rights Impact Assessment: A structured evaluation of the actual or potential impacts of platform policies, technologies, or laws on rights such as privacy and freedom of expression (BSR, 2013).

Imposter content: False information that impersonates a genuine person, organization, or trusted source (First Draft, 2017).

Incitement to violence: Language that intentionally encourages or advocates physical harm against individuals or groups (United Nations, 1966).

Independent oversight: A governance mechanism free from both government control and platform profit motives to ensure fair and transparent accountability (Meta Oversight Board, n.d.; UNESCO, 2023).

Internet of trust: A UNESCO concept describing a digital ecosystem where information is credible, platforms are accountable, and human rights are central to governance (UNESCO, 2023).

Malinformation: Information that is based on real facts, but manipulated to inflict harm on a person, organization or country (UNDP, n. d.)

Misinformation: Information that is false, but not created with the intention of causing harm (UNDP, n. d.)

Pre-bunking: A proactive strategy of warning users in advance about misleading narratives or manipulation tactics they may encounter (van der Linden, 2022).

Right to appeal: The right of a user to challenge a platform's moderation decision, such as content removal or account suspension, through a fair and transparent process (Santa Clara Principles, 2021).

Safety-by-design: An approach where user safety, rights, and risk mitigation are built into products and systems from the earliest stages of development (Australian eSafety Commissioner, n.d.).

Synthetic media / deepfakes: Audio, video, or images generated or significantly altered using artificial intelligence to make someone appear to say or do something they did not (Chesney & Citron, 2019).

User grievance mechanism: A formal, accessible, and transparent channel through which users can complain about platform decisions and receive timely remedies (United Nations, 2011).

Watchdogging: The act of civil society organizations or media independently monitoring platform practices and government regulations to hold them accountable to public interest standards (UNESCO, 2023).

References

Acharya, U. & Shrestha, U. (2026). *In the absence of truth, misinformation becomes harmful: Nepal's experience shows why fact-checking matters in crises. Poynter / IFCN.* https://www.poynter.org/ifcn/2026/in-the-absence-of-truth-misinformation-becomes-harmful-nepals-experience-shows-why-fact-checking-matters-in-crises/

Acharya, U. (2023). Promoting digital literacy with scarce resources. In Worthington, J. (Ed.), *Managing the misinformation effect — The state of fact-checking in Asia.* International Federation of Journalists. https://medium.com/disinformation-in-nepal/the-state-of-fact-checking-in-asia-nepal-c5a83c1b82

Acharya, U., & Shrestha, U. (2026). *Mithya suchana ra tathyajanch: Hatekitaab* [Misinformation and fact-checking: A handbook]. Center for Media Research – Nepal.

Adhikari, D. (2024, December 10). Inside Nepal's battle against misinformation and disinformation. *CSO Hate.* https://www.csohate.org/2024/12/10/nepal-misinformation-disinformation/

Africa Check (n.d.). *How we fact-check.* https://africacheck.org/how-we-fact-check

Amnesty International. (2021). *Sri Lanka: Writer detained for short story.*

Amnesty International. (2026). Bangladesh: Meta's content moderation delays risk fueling real-world violence. 16 March 2026. https://www.amnesty.org/en/latest/news/2026/03/bangladesh-metas-content-moderation-delays-risk-fuelling-real-world-violence/

Article 19. (2020). *International standards on freedom of expression and access to information.*

Associated Press. (2025, September 4). Nepal blocks Facebook, X, YouTube and others for failing to register with the government. https://apnews.com/article/3b42bbbd07bc9b97acb4df09d42029d5

Balasuriya, N. (2025). *Supreme Court ruling on Galle Face protest case.*

Bangladesh NGOs Network for Radio and Communication. (n.d.). *Reports on community media and digital inclusion.*

BBC. (2012). Bangladesh rampage over Facebook Koran image. 30 September 2012. https://www.bbc.com/news/world-asia-19780692

Business Standard. (2019). From Ramu to Borhanuddin: The same playbook used in inciting communal violence. 22 October 2019. https://www.tbsnews.net/bangladesh/ramu-borhanuddin-same-playbook-used-inciting-communal-violence

Centre for Policy Alternatives (CPA): *Liking violence: A study of hate speech on Facebook in Sri Lanka.* https://www.cpalanka.org/liking-violence-a-study-of-hate-speech-on-facebook-in-sri-lanka/

Centre for Policy Alternatives. (n.d.). *Reports on digital governance and media freedom.*
Democracy Reporting International (DRI): *Analysis of the Discourse on Facebook in Sri Lanka.* https://democracy-reporting.org/en/office/global/publications/social-media-monitoring-analysis-of-the-discourse-on-facebook-before-and-after-sri-lankas-2019-presidential-election

Dhaka Tribune. (2025, October 1). *32 % children in Bangladesh victims of online violence, activists warn.* https://www.dhakatribune.com/bangladesh/dhaka/392990/32%25-children-in-bangladesh-victims-of-online

Dhungana, S. (2020). Human dignity and cross-border migrants in the era of the COVID-19 pandemic. *Social Sciences & Humanities Open.* https://pmc.ncbi.nlm.nih.gov/articles/PMC7455169/

Digital Rights Nepal. (n.d.). *Policy briefs on internet governance and digital rights.*

Freedom Forum (Digital Freedom Coalition). (2025). *DFC condemns state's barbaric suppression of GenZ protest.* https://freedomforum.org.np/dfc-condemns-states-barbaric-suppression-of-genz-protest/

Freedom Forum. (2025). *Digital Rights Conference 2025: Ethical governance and protection of digital rights in Nepal.* https://freedomforum.org.np/news/detail/825

Freedom Forum. (2025). *FF joins #KeepItOn coalition against government ban on Telegram app.* https://freedomforum.org.np/news/detail/804

Global Freedom of Expression. (2025). *Case analysis: Nathasha Edirisooriya.*

Government of Bangladesh. (1972). *Constitution of the People's Republic of Bangladesh.*

Government of Bangladesh. (2009). *Right to information act, 2009.*

Government of Bangladesh. (2018). *Digital security act, 2018.*

Government of Nepal. (2007). *Right to information act, 2007.*

Government of Nepal. (2008). Electronic transactions act, 2008.

Government of Nepal. (2015). Constitution of Nepal.

Government of Sri Lanka. (1978). *Constitution of the Democratic Socialist Republic of Sri Lanka.*

Government of Sri Lanka. (2016). *Right to information act, No. 12 of 2016.*

Hashtag Generation*: Social Media Monitoring of the 2019 Presidential Election.* https://hashtaggeneration.org/wp-content/uploads/2020/01/social-media-monitoring-report.pdf

International Fact-Checking Network. (n.d.). *IFCN code of principles. The Poynter Institute.* https://ifcncodeofprinciples.poynter.org/

Journalism Authority. (n.d.). *Fact-checking and verification in journalism.* https://journalismauthority.com/fact-checking-and-verification-in-journalism?utm_source=chatgpt.com

Kathmandu Post. (2020, April 1). *Fear and loathing: Stigma and ostracisation for those who've returned from abroad.* https://kathmandupost.com/national/2020/04/01/fear-and-loathing-stigma-and-ostracisation-for-those-who-ve-returned-from-abroad

Lamichhane, P. (2024, May 31). *False information ecosystem. The Rising Nepal Daily.* https://risingnepaldaily.com/news/43775

Macdonald-Kelce Library. (n.d.). *How to fact check like a pro.* https://utopia.ut.edu/FakeNews/factcheck

National Human Rights Commission of Nepal. (2020). *Rights of Migrant Workers in the Clutches of COVID-19 Pandemic.* NHRC Nepal.

Nepal News. (2025, February 13). Internet abuse cases on rise. https://english.nepalnews.com/s/society/internet-abuse-cases-on-rise/

Nepalkhabar. (2025, July 30). Civil society urges govt to lift ban on Telegram. https://en.nepalkhabar.com/news/detail/14807/

NewAge. (2025). Prothom Alo, Daily Star offices attacked, set on fire. 19 December 2025. https://www.newagebd.net/post/country/285518/prothom-alo-daily-star-offices-attacked-set-on-fire

Office of the United Nations High Commissioner for Human Rights. (2011). *Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework. United Nations.* https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

OHCHR. (2011). *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework.* [Foundational for Unit VI user grievance mechanisms]. Available at: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

OHCHR. (2020). *The B-Tech Project: Bridging the gap between human rights and technology (Guidance on Governance and Accountability). [Relevant to Unit VI].* Available at: https://www.ohchr.org/en/instruments-and-mechanisms/b-tech-project

PEN International. (2025). *Sri Lanka drops charges against writer Shakthika Sathkumara.*

Potter, W. J. (2022). *Media literacy* (10th ed.). SAGE Publications.

Regmi, P., Dhakal Adhikari, S., Aryal, N., Wasti, S. P., & van Teijlingen, E. (2022). Fear, stigma and othering: The impact of COVID-19 rumours on returnee migrants and Muslim populations of Nepal. I*nternational Journal of Environmental Research and Public Health*, 19(15), Article 8986. https://doi.org/10.3390/ijerph19158986

Reuters. (2025, September 9). *Nepal lifts social media ban after protests leave 19 dead, minister says.* https://www.reuters.com/world/asia-pacific/nepal-lifts-social-media-ban-after-protests-leave-19-dead-minister-says-2025-09-09/

Spector, J. (2020). *UN Working Group on Arbitrary Detention opinion on Sri Lanka.*

Substance Abuse and Mental Health Services Administration. (2014). *SAMHSA's concept of trauma and guidance for a trauma-informed approach.* HHS Publication No. (SMA) 14-4884. *(Relevant to the "trauma-informed approaches" for helplines mentioned on pages 10-11).*

Supreme Court of Sri Lanka. (2022). *SCFR 246/2022 and related cases.*

The Truth in Journalism Fact-Checking Guide (TiJ). (n.d.). *How to fact check.* https://thetijproject.ca/guide/how-to-fact-check/

UN Women. (2020, July 11). *In Nepal, a woman-managed quarantine centre tends to women's unique needs and recovery.* https://www.unwomen.org/en/news/stories/2020/7/feature-nepal-woman-managed-quarantine-centre

UNESCO. (2023). *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multi-stakeholder approach.* https://unesdoc.unesco.org/ark:/48223/pf0000387339

UNICEF Bangladesh. (2025, February 11). *UNICEF youth poll: Misinformation is the leading cause of stress for youth on social media.* https://www.unb.com.bd/category/bangladesh/misinformation-now-leading-cause-of-stress-for-youth-on-social-media-unicef/153279

UNICEF. (2020). *Child Safety Online: A practical guide for policymakers and stakeholders.* [Relevant to regional concerns about youth safety]. Available at: https://www.unicef.org/media/66681/file/Child-Safety-Online-2020.pdf

UNICEF. (2021). *Industry Toolkit: Children's Online Protection.* [Relevant to platform accountability]. *A*vailable at: https://www.unicef.org/media/94441/file/Industry-Toolkit-Childrens-Online-Protection.pdf

United Nations. (1948). *Universal declaration of human rights.*

United Nations. (1948). *Universal declaration of human rights.* https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (1966). *International covenant on civil and political rights.*

Wardle, C. (2019). *First Draft's essential guide to understanding information disorder.* First Draft News. https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Glossary.pdf

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe.* https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Wickramasinghe, R. (2023). *Colombo High Court bail decision in Edirisooriya case.*

World Health Organization. (2020). *Reporting on violence against women and girls: A self-learning guide for journalists.* https://www.who.int/publications/i/item/9789240015418

Center for Technology and Society. (2014). *The Marco Civil da Internet: Brazil's landmark internet bill of rights*. Fundação Getulio Vargas. https://www.fgv.br/cts/projects/marcocivil

European Digital Rights. (2022). *The Digital Services Act: Civil society's guide to a safer internet*. https://edri.org/our-work/digital-services-act-guide

Media Monitoring Africa. (2021). *Real411: A platform for reporting digital harms*. https://www.real411.org/about

Ranking Digital Rights. (2022). *The 2022 Ranking Digital Rights corporate accountability index*. New America. https://rankingdigitalrights.org/index2022

Ressa, M. (2022). *How to stand up to a dictator: The fight for our future*. Harper. (Or see: Rappler. (2020). *Mapping the disinformation ecosystem in the Philippines*. https://www.rappler.com/technology/investigations-disinformation-networks)

Bangladesh Telecommunication Regulatory Commission. (2023). Annual report 2022-2023. https://www.btrc.gov.bd/site/view/annual_reports

UN Women. (2025). *FAQs: Digital abuse, trolling, stalking, and other forms of technology-facilitated violence against women and girls.* https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women#:~:text=Governments%20and%20the%20technology%20sector,sharing%20someone's%20personal%20information%20online.

UNICEF. (2026). *Cyberbullying: What is it and how to stop it.* https://www.unicef.org/stories/how-to-stop-cyberbullying#:~:text=1.,seem%20like%20there's%20no%20escape.

Australian eSafety Commissioner. (n.d.). *Safety by design*. https://www.esafety.gov.au

BSR. (2013). *Conducting human rights impact assessments: Guidance and toolbox*. Business for Social Responsibility. https://www.bsr.org

Campbell-Stephens, R. (2020). *Educational leadership and the global majority: Decolonising narratives*. Routledge.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

European Union. (2022). *Regulation (EU) 2022/2065 on a single market for digital services (Digital Services Act).* Official Journal of the European Union. https://eur-lex.europa.eu

First Draft. (2017). *Understanding information disorder*. https://firstdraftnews.org

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

International Fact-Checking Network. (n.d.). *Code of principles*. Poynter Institute. https://www.poynter.org/ifcn/

Meta. (n.d.). *Coordinated inauthentic behaviour policy*. Meta Transparency Center. https://transparency.fb.com

Meta. (n.d.). *Meta ad library*. https://www.facebook.com/ads/library/

Meta Oversight Board. (n.d.). *About the Oversight Board*. https://www.oversightboard.com

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Santa Clara Principles. (2021). *The Santa Clara principles on transparency and accountability in content moderation*. https://santaclaraprinciples.org

UNDP. (n.d.). *Rise above: Countering misinformation and disinformation in the crisis setting*. https://www.undp.org/eurasia/dis/misinformation

United Nations. (1966). *International covenant on civil and political rights*. https://www.ohchr.org

United Nations. (2011). *Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework*. United Nations Human Rights Office. https://www.ohchr.org

United Nations. (2019). *United Nations strategy and plan of action on hate speech*. https://www.un.org

UNESCO. (2023). *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multistakeholder approach*. UNESCO. https://unesdoc.unesco.org

UNESCO. (2026). *Protecting critical voices: Guidance for human rights impact assessment on digital platforms*. https://www.unesco.org/en/articles/protecting-critical-voices

van der Linden, S. (2022). *Foolproof: Why misinformation infects our minds and how to build immunity*. Harper Collins.

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe. https://rm.coe.int

NOTES

NOTES

A Toolkit for Civil Society on Digital Platform Governance

Published under the initiative **"Safeguarding Freedom of Expression and Access to Information through the Implementation of the UNESCO Guidelines for the Governance of Digital Platforms,"** jointly undertaken by Tribhuvan University, Nepal; Jahangirnagar University, Bangladesh; and the University of Vocational Technology, Sri Lanka, with the support of UNESCO.

Published by:

Central Department of Journalism and Mass Communication (CDJMC), Tribhuvan University, with the support of UNESCO

unesco
United Nations
Educational, Scientific
and Cultural Organization

UNIVOTEC