

---

# ACCOUNTABILITY REGIMES FOR SOCIAL NETWORKS AND THEIR USERS

---

POLICY FRAMEWORK

SEPTEMBER 2022



Forum on  
Information  
& Democracy

---

# CONTENTS

---

About the Forum on Information and Democracy	3
About the Working Group	4
Foreword by Kjersti Løken Stavrum and Damian Tambini	5
Executive Summary	7
Definitions	9
Introduction: Shifting from Immunity Regimes to Accountability Regimes	11
1/ Different forms of regulation	11
2/ Platforms' decisions on accounts	12
3/ What is missing	13
<hr/>	
<b>CHAPTER 1: Accountability of Social Networks</b>	14
1.1 Rights of social networks	15
1.2 Obligations of social networks	15
1.3 Sanctions	21
<b>CHAPTER 2: General Accountability Regime</b>	22
2.1 Rights of account holders	23
2.2 Obligations of account holders	25
2.3 Sanctions	26
<b>CHAPTER 3: Special Regimes</b>	28
3.1 News media and journalists	29
3.2 Minors	30
3.3 Influencers	30
3.4 Highly influential accounts	31
3.5 Group administrators	32
<b>CHAPTER 4: Governance</b>	34
4.1 National level	35
4.2 International level	39
<hr/>	
Acknowledgments	44
Bibliography	47

# ABOUT THE FORUM ON INFORMATION AND DEMOCRACY

Providing democratic safeguards for the global communication and information space



## Commission on Information and Democracy

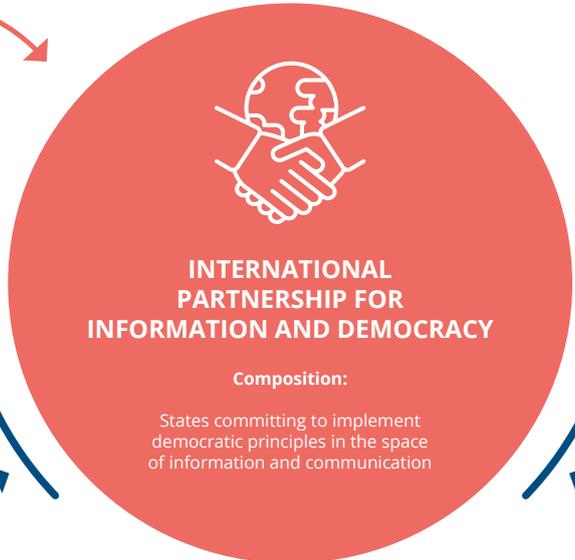
**Composition:**  
Nobel Prize laureates (peace, economics, literature), tech experts, journalists.

**Mission:**  
- Published the International Declaration which defines the universal principles for Information and Democracy and inspires the Partnership



## Annual Summits for Information and Democracy

**Objectives:**  
- Strengthening international cooperation  
- Working towards the implementation of the Forum's recommendations  
- Promoting dialogue between governments and the civil society



**EVALUATIONS**

**PUBLISHES** an assessment report ahead of the Summits of the Partnership

**EVALUATES** the norms, structures and architectures of the information and communication space



## International Observatory on Information and Democracy

**Missions:**  
- Providing states and society as a whole with periodic evaluations of the information and communication space

## IMPLEMENTING ENTITY



**RECOMMENDATIONS**

**DEVELOPS** regulation frameworks addressed to signatory States

**GATHERS** contributions and recommendations from experts



## International working group

**Assignments:**  
- Gathering international contributions and expertises  
- Suggesting recommendations to the states and stakeholders

**MOBILIZES**



## Civil society organizations and academia

**Missions:**  
- Contributing to develop recommendations and to the evaluation of the information space  
- Promoting the implementation of democratic safeguards in this space

# ABOUT THE **WORKING GROUP**

The working group on Accountability Regimes for Social Networks and their Users was announced during the Summit for Information and Democracy held on the margins of the UN General Assembly on September 21<sup>st</sup>, 2021.

It is composed of a Steering Committee and a team of rapporteurs. The Steering Committee has offered guidance to the rapporteurs and to the Forum.

A global call for contributions has allowed the working group to gather input from different regions and disciplines.

## **STEERING COMMITTEE MEMBERS:**

- **Kjersti Løken Stavrum**, Co-chair, CEO of Tinius Trust and president of PEN Norway
- **Damian Tambini**, Co-chair, Distinguished Policy Fellow in the Department of Media and Communications at the London School of Economics
- **Damian Collins**, Member of parliament, United Kingdom
- **Timothy Garton Ash**, Professor of European studies, Oxford University
- **Bernd Holznagel**, Director of the Institute for Information, Telecommunications and Media Law, University of Muenster
- **Anette Novak**, CEO, Swedish Film Institute, Former DG, Swedish Media Council
- **Jacob Mchangama**, Director, Justitia
- **Olivier de Frouville**, Professor of international law, Paris-Panthéon-Assas University
- **Richard Malka**, Lawyer specialized in freedom of expression
- **Karuna Nundy**, Lawyer, Supreme Court of India

## **RAPPORTEURS:**

- **Sarah Jamal**, lecturer in public law, Paris Panthéon Assas University
- **Anuchika Stanislaus**, tech policy analyst

# FOREWORD

## DEMOCRACY IS IN CRISIS, AND FREEDOM OF SPEECH IS BEING UNDERMINED ALL OVER THE WORLD.

*By Kjersti Løken Stavrum and Damian Tambini*

*Co-chairs of the working group*

The democratic idea – that the legitimacy of authority should be based on free deliberation between equal citizens – faces existential challenges of war, climate crisis and pandemic, amid deepening inequality and injustice. Updating media regulation cannot fix democracy or guarantee free speech in the face of these challenges and the authoritarians that seek to exploit them. But deep reform of social media is a necessary – and increasingly urgent – part of democratic renewal.

The radical transformation of information systems by the internet was expected to favor freedom of speech and a transparent society, enabling bottom-up social organization and a new plural, decentralized public discourse. This vision may yet come to pass. But recent years have shown that it will not do so without policy change as well as new thinking about free speech. Without new rules and institutions, the existing social media business model and governance – or lack of it – will continue to favor authoritarianism and populism. Data-driven, targeted and personalized media, based not on news values and fact checking but purely on engagement must be reformed, but our zeal for reform must not be used to justify a crackdown on free speech.

The political will for change comes from citizens. We have witnessed the threat these new internet gatekeepers pose – by undermining trusted free media; by fragmenting and coarsening public discourse and by undermining the shared factual basis of social trust and a sustainable civil society. In the short history of the internet, successive attempts have been made to strengthen it as a contribution to societal improvement: by developing ‘voluntary’ ethical codes, by encouraging social responsibility over profit, and imposing a range of new rules on content moderation. But these attempts have been largely ineffective. Where they did lead to change, the medicine was sometimes worse than the illness, as new forms of governance were used to stifle speech and capture free media.

The first paradigm of platform accountability grew out of competition between states as they sought to maximize the economic benefits of the internet. The next generation will be based on the need to protect democracy and human rights such as freedom of expression. In the past decade, immensely powerful gatekeepers have emerged on the internet, and civic societies need to agree on new approaches to ensure that the gatekeeping power of these new institutions is deployed in the public interest, and not for private or obscure interests. These gatekeepers occupy a critical position in democracy: their moderation increasingly operates as a private adjudicator of fundamental rights such as freedom of expression, but they are increasingly called upon by the state to moderate and regulate democratic debate. This is why the gatekeepers are entering regulatory pacts with governments all over the world, which is a moment both of opportunity and of danger for free speech and democracy.

Real reform will take time. The development of a new culture for online awareness as well as balanced regulation will not be quick or easy. It is global, multi-levelled and requires complex coordination among multiple stakeholders. And not least: reforms that improve democracies might have the opposite effect in totalitarian regimes.

The hard work of reforming media and information systems has begun. Many countries have published, and even passed, new laws, codes and regulations that aim to radically reform the incentives for social media and other internet intermediaries. But because of the dangers to speech freedom, this process of experimentation and institution building cannot be developed only by one or a few countries, it must be developed through genuine multilateral partnership, global standards, and global institutions.

Above all, successful governance reform requires clear international agreement on principles. The recommendations in this report offer a clear and decisive step in the direction of updating free media systems and reconciling them with a new, pro-democratic social media ecology. All states seeking to ensure the freedom of speech should urgently seek to endorse them.

# EXECUTIVE SUMMARY

## FROM IMMUNITY REGIMES TO ACCOUNTABILITY REGIMES

- > In recent years, the information and communication space has been drastically restructured by digital platforms and social media accounts.
- > No legal provision specifies how platforms should handle accounts, giving tech companies full flexibility to define the rights, duties, and sanctions of these accounts.
- > With their self-appointed and self-administered standards and unilateral decisions to sanction accounts, platforms hold the powers to decide who can or cannot participate in the public debate within their services.
- > Given the role these platforms have taken in structuring the public debate, these types of decisions should be framed by democratic institutions and following democratic principles. In short, there is a need to establish accountability regimes.
- > This accountability towards democratic authorities must be settled within a new national and international governance framework.

This report develops accountability regimes for digital platforms and their users. The recommendations presented below are the result of an international call for contributions, and a set of interviews and research carried out by the group of rapporteurs under the guidance of a Steering Committee. The recommendations outlined are addressed to states, social media platforms, civil society organizations, and social media users.

**Accountability regimes recommended in this report must be understood within the framework of international human rights law, in particular Articles 19 of the Universal Declaration of Human Rights and of the International Covenant on Civil and Political Rights. No recommendation in this report may be interpreted or used as a pretext for any state, group or person to undermine or destroy human rights and freedoms.**

### General accountability regime

#### Platforms should:

- > Create and enforce a general accountability regime for social networks users that would fix in law the rights, duties, and applicable sanctions of account holders;
- > Detail, as part of the general accountability regime, which type of content (e.g., disinformation, misinformation, hate speech) and practices (such as mass-reporting, astroturfing, inauthentic coordinated behavior) should entail the liability of account holders.

#### States should:

- > Ensure that sanctions on social networks (limiting the reach, suspending or terminating of an account, for instance) are made in accordance with human rights, i.e., the sanction must have a legal basis, pursue a legitimate aim, and be proportionate;
- > Implement an appeal mechanism for account holders to challenge decisions from platforms, by reference to a judicial authority.

## Special regimes

### **Platforms should:**

- > Create special regimes for news media and journalists, influencers, highly influential accounts, and group administrators, given their important role in shaping the public debate and their capacity to influence;
- > Create a special regime for minors;
- > Create a code of conduct for influencers, highly influential accounts, and group administrators to frame their activity;
- > Impose sanctions that reflect a user's level of influence and capacity to cause harm.

## Accountability of social networks

### **Platforms should:**

- > Create internal appeal processes enabling users to challenge a content moderation process or any sanction taken by the platform;
- > Have the right to refer specific cases to the national or transnational regulatory body to seek advice and provide transparency reports on the operation of these schemes.

### **States should:**

- > Hold platforms responsible in proportion to their influence, and impact of their contribution;
- > Hold platforms responsible for respecting national and international law. International law should be preferred when it is more favorable to the account holders;
- > Impose on platforms a transparency obligation and notification requirement to account holders when implementing sanctions;
- > Hold platforms responsible for complying with decisions from authorities, and apply sanctions or reinstate accounts and content according to the decisions.

## Governance

### **States should:**

**At the national level, establish a new or strengthen an existing independent administrative authority.**

- > This authority should have the mandate to serve as support for users of digital services, oversee platforms, serve as a research cluster, and advise on public policies and regulations at the national level.
- > This authority should have the necessary means and resources to achieve its objectives. A specific taxation on large digital platforms should be used to finance this authority.
- > The decisions taken by the authority should be binding to social networks and entirely independent from the government's executive branch and private interests.

### **Create a transnational body. This body should:**

- > Have the mandate of monitoring the application of accountability regimes by states and platforms;
- > Include a dispute-settlement body based on arbitration principles;
- > Include an investigative assistance body in charge of collecting, consolidating and preserving information and evidence of violations of international humanitarian law and human rights violations and abuses.

# DEFINITIONS

## 1/ ENTITIES

**Social network** is a category of websites with unique URL profiles, where each profile account can write a public commentary on another profile (it may disappear over some period of time or upon removal), and each profile holder can traverse<sup>1</sup> the network through a series of connections.

In this report, “social network” and “social media” are used interchangeably.

**Online service providers:** as stated in the Declaration on Information and Democracy, are entities that help structure the information and communication space by creating the technical means, architecture, and standards for information and communication. This includes digital platforms (“platforms”), which are defined as: “online sites and services that: (a) host, organize, and circulate users’ shared content or social interactions, (b) without having produced or commissioned (the bulk of) that content, (c) built on an infrastructure, beneath that circulation of information, for processing data for customer service, advertising, and profit.”<sup>2</sup>

In this report, “platforms” and “online service providers” are used interchangeably.

The report does not address all platforms but only those that provide social media services whose activity contributes to the public debate in our democratic societies.

Online service providers / platforms should be distinguished from:

- > **Hosting services:** those services that consist of the storage of information provided by, and at the request of, a recipient of the service (such as web hosting or cloud services);
- > **Intermediary services:** a “mere conduit” service that consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network (e.g., internet service providers).

**Very large online platforms (VLOPs):** refers to platforms that meet a threshold of 10% of consumers in the regional market.

## 2/ TYPES OF CONTENT

**Disinformation:** Information that is false and deliberately created to harm a person, social group, organization or country.<sup>3</sup>

**Misinformation:** Information that is false but not created with the intention of causing harm.<sup>4</sup>

**Legal yet harmful:** Content that is legal yet deemed harmful by the platform.

**Rumors:** Claims which do not hinge on their inherent truth value, and whose power arises from social transmission itself.<sup>5</sup>

**Conspiracy theories:** The belief that a hidden group of powerful individuals exerts control over some aspect of society.<sup>6</sup>

1 Boyd, Danah M. & Nicole B. Ellison (2007).

2 Gillespie, Tarleton (2018).

3 UNESCO (2018).

4 Ibid.

5 Berinsky, A.J. (2015)

6 Vermeule, Adrian & Cass Sunstein (2009)

### 3/ CONTENT MODERATION AND CURATION

**Content moderation:** When online platforms screen and monitor user-generated content based on platform-specific rules and guidelines to determine whether to host or continue hosting a specific piece of content under their terms of service. These decisions include removal of content, permanently, temporarily, or by geographical area.<sup>7</sup>

**Content curation:** Decisions regarding the reach, prominence, or amplification of certain content. These could include boosting, invisibilization, or demoting.<sup>8</sup>

### 4/ REGULATION

**State regulation:** Any binding legal or regulatory instrument that local, national, or regional public institutions enact through their legislative processes.<sup>9</sup>

**Self-regulation:** Online platforms define what kind of content is acceptable when using their services, often by creating their own terms of service. Platforms carry out regulation primarily in two ways: through moderation or curation of content.<sup>10</sup>

**Co-regulation:** A system in which the general guidelines and expected results of platform policies are defined in a legal instrument, with input from multiple sectors, which must be applied directly by platforms taking into consideration local and regional context and in line with human rights principles. An appropriate body, with guarantees of independence and autonomy, should oversee the companies' application of these standards. Co-regulation should include civil society and could potentially exclude governments.

---

7 Access Now (2020).

8 Ibid.

9 Ibid.

10 Ibid.

# INTRODUCTION

## SHIFTING FROM IMMUNITY REGIMES TO ACCOUNTABILITY REGIMES

As stated in the International Declaration on Information and Democracy, the global communication and information space is a common good of humankind and of great democratic value. As such, its management should be based on democratic institutions and democratic principles.

Digital platforms structuring this space hold paramount power over the ability to participate in the democratic process, given that users utilize these interfaces to exercise their right to information and their freedom of expression and opinion. The power of shaping the public debate and democratic life must come with inseparable guarantees of accountability.

Yet, the current situation and regulation of the online space prove otherwise, specifically when it comes to the management of accounts by social networks. There is too little oversight and transparency. We must shift from immunity to accountability.

### 1/ DIFFERENT FORMS OF REGULATION

The international traditional law on freedom of expression has been developed to grant free speech while providing reasonable restrictions to what citizens can or cannot say. International standards have been formulated, notably in the Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which stipulates that these restrictions shall “only be such as are provided by law and are necessary: (a) for respect of the rights or reputations of others; (b) for the protection of national security or of public order (order public), or of public health or morals”.<sup>11</sup> International human rights law, and in particular Article 19 of the ICCPR should serve as a framework of first reference for systemic and individual content moderation decisions and practices. As such, the ICCPR should take precedence if conflicting with regional international human rights instruments.

This entails that content moderation practices must be legitimate, necessary and proportional within the framework of Article 19(3) ICCPR (restrictions on freedom of expression) which sets out the limitation grounds of freedom of expression. For hate speech, the applicable threshold should be based on Article 20(2) ICCPR (prohibition of advocacy of hatred) and take into consideration the Rabat Plan of Action’s six-part threshold test for context, speaker, intent, content and form, extent of dissemination, and likelihood of imminent harm before taking enforcement action. For disinformation, the grounds in Article 19(3) ICCPR and Article 25 ICCPR (right to participate in voting and elections), should be adhered to. Further vague, blanket policies for removal are incompatible with Article 19 and only disinformation entailing real and immediate harm should be subject to the most intrusive restrictive measures, such as content removal. In determining the limits of disinformation, the post’s content, its context, its impact, its likelihood of causing imminent harm, and the speaker’s intent must be decisive factors.

All democracies consider freedom of expression as the bedrock of their political regimes, even though there may be different cultural interpretations on how far restrictions on freedom of expression may be reasonable. National laws and regulations can take different forms within a shared democratic conception, from an absence of state intervention, where freedom of expression is considered the strongest and most necessary value, at the expense of others, to a more interventionist model in which the exercise of the state appears as an additional guarantee of effectiveness. The effectiveness of free speech in democratic countries relies on the reasonability of speech restrictions and their application under the rule of law.

<sup>11</sup> Office of the United Nations High Commissioner for Human Rights (1966).

The current laws on freedom of expression allow ruling on the legality or illegality of the content and, in relevant cases, defining what compensation should be made by the offender to the offended for the harm suffered. These laws apply to all participants in the public debate online and offline.

Media outlets have an historic role in shaping the public debate, and as such have been subjected to specific regulations (media law), both to ensure freedom of the press and to set basic requirements for information reliability and plurality.

Media law sometimes defines rights and more often obligations to news media and the public. News outlets and their editorial leadership are liable for the publication and can be sued for press offenses such as libel or breach of privacy. Traditionally, they follow self-regulatory ethical and professional standards<sup>12</sup> for respecting the truth and granting rights to the public, such as the right of reply or correction. Obligations to provide plural information to their audience are often required, especially in the audiovisual field. Along with these duties, news media have been given specific rights to serve the public interest and public fundamental rights, without any interference from public authorities or private interests.

In recent years, the information and communication space has been drastically restructured by digital platforms, which are not accountable for the content they host on their services. They have given rise to a new actor in the information and communication space: social media accounts. No legal provision specifies how platforms should handle accounts, giving tech companies full flexibility to define the rights, duties and sanctions of these accounts.

## 2/ PLATFORMS' DECISIONS ON ACCOUNTS

The rights, duties and sanctions on accounts are not framed by law, but solely by the terms and conditions established by platforms, mostly with commercially oriented purposes and without due diligence on their impact on human rights and democracy. Because of their international application regardless of borders and cultures, terms and conditions are not always consistent with international or national law.

In the event of a violation of their terms of service (ToS), platforms have taken sanctions on accounts ranging from reducing the reach of a specific account to permanently banning the user. These decisions have a major impact on the public debate, specifically when it concerns public figures, politicians, or opinion leaders, and are taken with no accountability to the public or any independent authority.

Platforms have also taken further action to provide context or background on accounts, verifying some and labeling others. Twitter's policy on government and state-affiliated media account labels, or Facebook's profile verification policy, illustrate these efforts of self-regulation to protect the integrity of public debate, sometimes providing accounts of public figures with specific moderation exemptions.<sup>13</sup>

But this self-regulation of accounts raises concerns about its opacity, arbitrariness and unilaterality. On the one hand, digital platforms are not accountable for applying their own terms and conditions, and sometimes grant privileges to certain account holders enabling them to breach the platform's own policies.<sup>14</sup> On the other hand, other account holders face unexplained and discretionary moderation decisions, even though they are publicly recognized as reliable and professional sources of information.<sup>15</sup> In the absence of due diligence on the impact of their internal rules and policies, platforms have adopted practices detrimental to the rights of their users.

12 For instance, the Declaration of the Duties and Rights of Journalists, also known as the Munich Charter, defines ten duties and five rights for the journalist profession. For more information, see: European Federation of Journalists (1971).

13 Twitter (2019).

14 Several examples could be mentioned, such as Facebook's XCheck internal programme, revealed by The Wall Street Journal in September 2021, and Meta's change in its Glorification of Violence Policy in some European and Central Asian countries. For more information, respectively see: Horwitz, Jeff (2021) and Vengattil, Munsif & Elizabeth Culliford (2022).

15 Just to name a few cases: the case of French journalist David Thomson, the account of The Kashmir Walla magazine in India, or the account of the Ecuadorian investigative news outlet Periodismo de Investigación. For more information, respectively see: Reporters Without Borders (2016), Reporters Without Borders (2021) and Observacom (2022).

The inability of account holders to properly appeal the sanctions enforced by platforms is also problematic. Firstly, shadow banning or other unnoticed sanctions prevent account holders from exercising their right to appeal, given their lack of knowledge. Secondly, the appeal process is often made under obscure internal mechanisms. Each company has developed its own appeal modalities and mechanisms, usually consisting of a review by the support team.<sup>16</sup> The most famous exception is Meta's Oversight Board, which is composed of independent experts. Since 2020, the Board has selected the most complex removal or suspension cases and decided in the second instance if the moderation was appropriate or should be overturned. Meta can also ask the Oversight Board for recommendations on difficult moderation issues, but these are non-binding.

With their self-standards and unilateral decisions to sanction accounts, platforms hold the powers to decide who can or cannot participate in the public debate taking part in their services. Given the role these platforms have taken in structuring the public debate, these types of decisions should be framed and scrutinized by democratic institutions, and following democratic principles.

### 3/ WHAT IS MISSING

Specific rights and obligations should be defined for account holders and digital platforms to establish a clear framework respecting human rights and the principles of rule of law. This democratic framework must establish necessary, proportionate and fair sanctions in case of violations of these rights and obligations, including breach of terms and conditions by account holders.

A due and transparent process of sanctioning must be developed to prevent platforms from moderating in an opaque and arbitrary manner. It must provide account holders a right to appeal to an internal dispute resolution mechanism set up by the platform, and inform them on how to refer to a national authority or court in case they want to appeal the internal decision.

This shift from immunity regimes to accountability regimes for platforms and account holders is essential to ensure the legitimate control of democratic institutions over the public debate. This accountability towards democratic authorities must be settled within a new governance framework. A national administrative authority should be mandated to make sure digital platforms respect all their obligations and the rights of their account holders. Users should easily have access to independent judicial institutions dedicated to litigation involving platforms.

The power that digital platforms hold on the public debate goes far beyond the national scale. To offset their tremendous influence on the worldwide space of information and communication, national institutions won't be enough: international bodies must be created or mandated to grant the respect of universal human rights by platforms and hold them accountable under international law.

Whereas the latest legislation on digital platforms was mainly focused on reducing the systemic risks that they pose to society,<sup>17</sup> no legislative initiative aims at fixing the problematic loopholes surrounding the accountability of platforms and users over accounts and their related content.

This report develops accountability regimes for digital platforms and their users. In line with the principles of the International Declaration and Partnership for Information and Democracy, it intends to create democratic safeguards protecting the exercise of freedom of opinion and expression and the right to information in the digital space of information and communication.

---

16 For more information on the policies of each platform, see the comparative analysis "How To Appeal." from Electronic Frontier Foundation (EFF), accessed on 18 Aug 2022 via: <https://www.onlinecensorship.org/en/pages/1214ccb8-e1d5-472a-91e8-745d895b4dff>

17 We can see this approach in the Digital Services Act (DSA) in the EU (see Appendix 1) or in the Online Safety Bill in the UK.

# Chapter 1:

# Accountability of Social Networks

It is essential that social networks respect the principles and guarantees that ensure the democratic nature of the information and communication space. Recalling the International Declaration on Information and Democracy, social networks must be responsible in proportion to their influence and the impact of their contribution.

# 1.1 RIGHTS OF SOCIAL NETWORKS

---

**Social networks should have the right to:**

## **Seek advice from a transnational regulatory body**

Platforms can appeal to the expertise of the transnational regulatory body when encountering a difficult situation.

## **Establish a definition for content that is deemed legal yet harmful**

Social networks have the right to include in their terms and conditions a definition for content that is considered legal yet harmful, and take appropriate<sup>18</sup> moderating decisions to mitigate its spread, if the following conditions are met:

- Such definition should be elaborated through an inclusive process of deliberation, with the participation of affected persons and communities and civil society organizations.
- Both the definition and the decisions taken on its basis must be in conformity with international human rights law, including Article 19 of the ICCPR.

# 1.2 OBLIGATIONS OF SOCIAL NETWORKS

---

**We recommend that social networks commit to the following obligations with respect to their relationship with the users, compliance with authorities, and decision-making practices.**

## **National and international law**

Social networks should respect their obligations under international<sup>19</sup> and domestic law consistent with international obligations of the state in the field of human rights and fundamental freedoms. In the spirit of Article 5, par. 2 of the ICCPR, social networks should apply the law that is the most favorable to human rights and specifically to the freedom of expression of the user who posted the content.

## **Fair and transparent decision process**

Social networks should respect the 5 Cs principle in their moderation solutions: customization, compatibility, community, care, and centralization-free.<sup>20</sup> Additionally, the moderation process should be necessary, proportionate, and conform to the principles in the terms of service.

**Platforms should analyze a series of parameters** before taking any decision to block, demonetize or promote content or accounts, whether that decision is made through automated moderation practices or by human moderators, by taking into account the:

- quality of the sender (greater caution should be exercised for journalists or persons having a great level of influence);
- intention of the author of the publication (is the image posted with an activist/informational purpose, etc.);
- format of the post and the context surrounding its content;
- likelihood of imminent harm to users and the public;
- extent of dissemination of the content;
- participation (or not) of the information in a debate of general interest in the society.

---

18 Sander, Barrie (2020).

19 United Nations (2011).

20 Ermoshina, Ksenia (2022).

### Automated content moderation

Social networks should explain the logic of the algorithm, and the criteria to which the algorithm responds. Algorithms should follow a **principle of explicability** towards the users and explain<sup>21</sup> the:

- number of times content has been curated, moderated and ordered;
- numbers broken down by action taken;
- criteria and data used to train the algorithms to identify, moderate, prioritize and personalize content;
- collection of data, including personal data;
- biases of the data and biases reproduced by the algorithms;
- potential misuses and abuses of the algorithms, false positive / false negative rates;
- procedures used to correct errors.

Platforms should set up teams dedicated to the quality of user dialogue through a mediation service to ensure that users understand the information correctly. Explanations of the algorithm should be provided in all languages in which the social network operates.

The platforms should organize a form of «reachability» of algorithmic systems, in particular by systematically identifying within each company or administration a team responsible for the operation of an algorithm as soon as it processes people's data. This team must be easy to contact and have the means to respond quickly to requests received. This could be achieved by:

- setting up teams dedicated to the quality of user dialogue;
- setting up a mediation service to ensure that users understand the information they need.

### Effective notice procedure

Social networks should provide every account holder with an effective notice procedure<sup>22</sup> that meets the principles of accessibility, clarity and readability. Users should be informed immediately if content is taken down, blocked, demoted or demonetized, or associated with a content warning.

Social networks should set up a supervised notice procedure to report objectionable content on platforms, such as those that make a call to terrorism. Additionally, the user who notified the content should be informed of the evolution of the request. The **notice procedure should be simplified**: it should be possible to complete the notice in a maximum of three clicks and it should be standardized on all platforms.

### Appeal process

In order to ensure that there is a fair hearing for users, social networks should commit to establishing appeals processes.<sup>23</sup> This includes an appeal body to contest the “shadow-ban” decision, and the establishment of a dialogue between users and the teams responsible for the operation of algorithms using personal data, who should be representative and trained on the impact of their activity.

### Special regimes

Social networks should recognize, as a minimum requirement, the special regimes identified in Chapter 3 and establish procedures to adapt to their additional rights and duties accordingly.

### Cooperation with other entities

Social networks should submit to the arbitral tribunal disputes concerning content moderation where the situation is of some complexity (a case concerning an international public person must be regarded as a complex case).

21 Tsamados, Andreas et al., (2022).

22 Council of Europe (2021).

23 Common, MacKenzie, (2019).

They should transfer to the international investigative mechanism, on their own initiative, any information that contributes to demonstrating the violation of human rights, or any illegal act that qualifies as a serious crime (genocide, crimes against humanity, war crimes), as well as any relevant information at the request of this investigative mechanism.

The platform should transmit the element identifying the user to an independent judicial authority in event of legal proceedings under certain conditions:

- The act must be illegal under domestic law that is consistent with international human rights law.
- The state must rank highly according to the transnational regulatory body's principles of justice and freedom of opinion and expression. If the transnational regulatory body is persuaded that human rights of users will be respected, the identity of users may be revealed to the authority.
- The identification of the user must be requested by an independent and impartial judicial authority in the context of a procedure respecting fair trial.

Competent **authorities should**:<sup>24</sup>

- have access to data identifying the persons who posted the problematic content;
- have access to the post concerned;
- set up an automatic transmission of an alert to a dedicated public platform, or even a platform accessible by international investigators.

### Compliance with authorities' decisions

Social networks should commit to comply with the decision of domestic authorities that are consistent with international human rights law.

Whenever an independent judicial authority deprives the account holder of access to the social network, **the platform has the obligation to apply the sanction.**

### Cooperation with the transnational regulatory body

Social networks should commit to participate within and cooperate with the transnational regulatory body (see Chapter 4).

### Appointment of a legal director

It is essential that social networks commit to appointing a legal director to represent them in the state in which they operate. The legal representative will be the main interlocutor with the competent national authorities and will represent platforms in legal proceedings.

### Algorithmic liability

Social networks should commit to being liable for the harm caused by their algorithms. Any social network using algorithms that create a toxic user experience, for example by amplifying hate, should be designated as content producers and held responsible for the harm caused.

### Compensation to victims

Social networks should commit to establishing a procedure that respects the principles of fair trial in order to determine the existence of damage, and the form and amount of compensation required for the victim.

They should have the ability to **provide compensation** if and when an error in moderation has imposed financial loss on the user.

---

24 Heldt, Amélie & Stephan Dreyer (2021).

### Terms and conditions

Social networks should publish their policies regarding what user-generated content and behavior is or is not permitted: for rules about content and targeting for advertising; for content moderation, content ranking, content targeting, and socializing recommendations; and for processing and disclosure of user data. Terms and conditions should be clear for users to access and understand, and should include applicable sanctions for account holders in the event of non-compliance.

Terms and conditions should align their definitions of problematic content, such as “hate speech” and “incitement to terrorism”, with those developed by public international institutions, such as relevant UN treaty bodies, special-procedure mandate holders and other experts, such as the WHO, and international courts and tribunals.<sup>25</sup>

### Amendment of terms and conditions

Social networks should set up a **public and transparent amendment procedure** of their terms and conditions and include appropriate transitional measures.

- If there is any update, users should be notified and have access to the previous versions of the charter, and the reasons why it has evolved.
- The terms and conditions should be easy to find on their sites upon registration and during general use of their platforms, and available in different languages in line with their global reach.

Terms and conditions should include a policy on collecting, preserving, and sharing content that may evidence violations of international law, allowing users to report on human rights violations during social unrest, and on restriction of internet access in a country.

### User participation

Social networks should establish a mechanism for users to amend the terms and conditions. This should include user representation in moderation governance, decision-making bodies, platform advisory boards (content advisory councils, Trust and Safety Council, etc.), as is the case in some companies for employees.

### Civil society inclusion

Social networks should include civil society in self-regulation<sup>26</sup> through a digital platform that could aggregate feedback and litigious cases by building appropriate regulation tools (indicators, algorithms, flagtaggers etc.). Civil society should be involved when the platform conducts its risk assessments, designs its risk mitigation measures, and draws up codes of conduct.

The modalities of civil society participation should respect a transparent procedure, show a representation of different categories (journalists, NGOs, etc.), and have an equal geographical and gender representation.

### Curation and moderation of content

Social networks must publish policies regarding content moderation, content ranking, content targeting, and socializing recommendations. They must maintain up-to-date reference documents on each core function of the algorithms, including ranking, targeting, moderation and social recommendations, as well as detection of content.<sup>27</sup>

### Detection of new forms of illegal or harmful speech

Social networks should invest in identifying complex online harmful communication, especially when extremist users employ generic euphemisms to evade easy detection.

25 Kaye, David, (2019).

26 Isaac, Henri & Louis-Victor de Franssu (2022).

27 Forum on Information & Democracy (2020).

### Human and automated moderators

Social networks should set up a transparent ratio between human and automated moderators. The removal of content should be treated by human moderators and not algorithms, according to an established threshold of content complexity. The threshold should be determined by a coalition of experts in international human rights laws and civil society. They should be involved in the design and evaluation of the degree of complexity.

### Human moderation

Social networks should be obliged **to ensure a certain level of human moderation for each country, which should be representative** of the population, so that platforms can make decisions according to the context:

- Moderators should represent minorities.
- Moderators should be regularly trained in the local geopolitical context, applicable local law and international law.
- Human moderators should be fluent in the vernacular language used in the posts they treat.

Moderators should be psychologically supported and have suitable working conditions.

### Access to researchers and civil society organizations

Social networks should provide researchers and civil society organizations with access to **contextualized information** that surrounds aggregated data currently provided for all moderated content.

This transparency obligation will bring any kind of bias or patterns of discrimination to light, and help protect minorities or disproportionately over-moderated<sup>28</sup> regions of the world. Researchers and organizations can help platforms discern whether content is illegal or should be removed. Researchers and civil society organizations may also support the victim-users during the proceedings. It allows legislators to adopt laws and regulation to respond to realities through exhaustive in-the-field data.

These data sheets should be anonymized.

Social networks should open **algorithmic systems (source codes)** and make them accessible to independent researchers.

Social networks should also implement **testing procedures by making users' activity data available** via an API (application programming interface), which could be the basis for collective actions based on their data (data unionization).

### Disclosure requirements

In light of the fact that data platforms already share<sup>29</sup> information regarding online terrorist and violent extremist content, they should extend transparency reporting to all other categories of content. All additional disclosure requirements should be audited by an independent auditor, as is the case with corporate financial reporting.

Disclosures should draw guidance from the Santa Clara Principles on Transparency and Accountability in Content Moderation,<sup>30</sup> established by civil society organizations, which set out the minimum level of detail that platforms should be expected to disclose.

### Impact assessment and audits

Social networks should commit to publishing annual impact assessments on algorithms in accordance with responsibilities to conduct such audits, as laid out by different states.

28 Pen America (2021).

29 OCDE (2021).

30 The Santa Clara Principles (2018).

Platforms should establish a regular and public systemic risks assessment to ensure the compliance of algorithms with their initial objectives, and to identify potential damaging biases. Algorithms should be tested regularly, using human moderators.

### **Inclusive recruitment**

Social networks should encourage recruitment that promotes cultural, social and gender diversity for every profession involved in algorithm design, and be transparent as to diversity among algorithm designers.

### **Training**

Social networks should provide ongoing training for algorithm designers (developers, programmers, coders, data scientists, engineers) to encourage awareness of the implications of their activity and their duty of care.

### **Content prioritization parameters**

Social networks should leave the content prioritization parameters accessible and editable to users. This should include:

- transparency on “shadow-ban” policy: the reasons, the amount of content that has been subject to this technique, etc. Also, transparency as regards “filter bubble” bias, and the availability of data used by the algorithm to select the next content the user will see;
- informing the user that their content is being deprioritized.

### **Mass-reporting**

Social networks should commit to prohibiting mass-reporting in moderation rules, and place limits on how many reports any single account can make in a day.

### **Illegal content or behavior**

Social networks should create a portal available to all users in order to report illegal content or behavior. The portal would assist users in identifying who is responsible for the content, and may lead to action by the legal authorities.

### **Automated accounts**

Social networks should require users to undergo a Captcha test on a regular basis (e.g., once a month), and not just at initial log-in, in order to limit the action of automated accounts. Accounts operated by bots will be easily tracked and subject to disconnection or other sanctions.

### **Detection of artificial virality**

Social networks should establish the necessary means to better detect artificial virality of content. Platforms should exercise their ability to add a friction mechanism to limit the virality of content.

### **Reappearance of notified illicit content**

Social networks should technically ensure that content that has been notified and removed is not able to reappear.

### **Reliable and comprehensive information**

Social networks should prioritize through their algorithms the delivery of information in accordance with a methodology that ensures reliability and independence, such as an ISO standard (for example, the Journalism Trust Initiative [JTI]). Large platforms which participate in the structuring of the information and communication space must respect the principle of political, ideological and religious neutrality. Algorithms that disseminate and process information and ideas must be neutral with regard to the

interests of those who control them, with the exception of political and social-issue-related advertising, which must be explicitly identified. Indeed, political and social-issue-related advertising should be presented to users in the context of different points of view.

#### Users' information on changes to the algorithm

Social networks should update users about changes to their algorithms, especially when the change has an impact on account-holder financial sustainability (such as for news media and influencers).

## 1.3 SANCTIONS

**It is necessary that the sanctions adopted against social networks are adapted to the seriousness of the breach committed and to the financial capacity of the network. The sanction should be issued either by the national independent authority or an independent judicial authority.**

#### Sanctions must reflect the platform's level of influence and resources

The decision on sanctions must take into account the seriousness of the non-compliance with the obligation, as well as the financial capacities of the platform.

#### Range of sanctions

Sanctions could range **from a public alert to a civil fine that could amount to 6% of a platform's annual global turnover.**



#### RECOMMENDATIONS TO STATES

- > Ensure that social networks meet their obligations.
- > Impose sanctions in cases of non-compliance.

# Chapter 2: General Accountability Regime

It is important to recognize that account holders, as new participants in the global information and communication space, have rights to protect and responsibilities to adhere to. Account holders too often act with impunity hidden behind pseudonyms. It is essential to recognize that they are responsible for the content they disseminate or contribute to disseminating, and that they can be held liable for the harm caused. The following sections present the rights and obligations attributed to general account holders, as well as the sanctions and remedies applicable.

---

## 2.1 RIGHTS OF ACCOUNT HOLDERS

---

The following delineates the rights that users should have when engaging with social networks.

### Rules governing speech

Users have the right to be informed of the rules governing speech that they are expected to abide by online. Account holders have the right to have clear and exhaustive information on a platform's moderation policy.

- Information should include a list of clearly objectionable content which will incur sanctions. This list should include all the current patterns, but also those that fall through the cracks, such as child abuse images, and Meta's «XCheck» program, among others.
- All users should know the criteria on which a platform's moderation decisions are based, who the policy applies to, what criteria will be used to define the public interest, what adjudication is being made and in which circumstances, and possible sanctions.

### Demotion and removal of content

Users have a right to be informed if and when their content is subject to moderation or removal ('invisibilization').<sup>31</sup> The information must include the reasons for the decision as well as the avenues available to challenge it.

Notifications should be detailed enough to allow the user to specifically identify the restricted content, and should include information on how the content or account was detected, how it was evaluated, and why it was deleted or restricted. However, in some cases, an independent and impartial judge may order that a user who, for example, has repeatedly demonstrated harassment activities towards another user must not be informed of the invisibilization of the content they have posted, so as to safeguard the safety of the victim(s) during the proceedings against the perpetrator.

### Appeal

Account holders have a right to appeal a social network's moderation decisions, such as the removal of a post or the deplatforming of an account. If the social network's appeal mechanism is unsatisfactory, users have a right to appeal to an independent administrative authority or judicial authorities.

### Access to fair trial

Account holders have the right to have a fair trial procedure, and the right to access an independent administrative authority or judicial authorities to challenge moderation decisions and sanctions imposed on them by social networks.

### Access to reparations

Victims have the right to reparation for the harm suffered online and must be able to take part in the reparations process. Reparations should compensate for all forms of damage, material or moral. The harm must be compensated for in its entirety, without the amount of compensation exceeding the amount of damage suffered.

Reparations should prioritize financial compensation, except when a symbolic measure can contribute

---

31 Nicholas, Gabriel, (2022).

to repairing the damage suffered. Thus, reparation can also take the form of a public apology published on the social network.

### Access to relevant general data

Users should be given the right to access relevant general data about moderation of content on the social networks, such as:

- content that was missed by the AI and human moderators, not reported, or not removed but should have been moderated;
- procedures put in place to minimize moderation errors;
- the number of pieces of content restored during the reporting period;
- automated content moderation procedures, and information on the purpose, accuracy indicators and safeguards applied;
- total number of moderation decisions applied by social networks.

### Autonomy and proactivity

Users should have the right to autonomy, and to be proactively part of the moderation processes.

- Users should be given adequate tools to moderate the content posted on their own account, page or group.
- Users should be active in deactivating, deleting or filtering comments under their publications, and be alerted when malicious content is reported on their space, so that they can moderate it themselves by applying their own specific rules and automated blocking plugins.

Users should be enabled to intervene through their representatives in the moderation process.

- Before the social networks' moderators are notified about content, it should be sent to a group of representative users who could take decisions in a collaborative manner, or individually through self-moderation of their own pages.

### Participation in the amendment of terms and conditions

Users should be able to participate and to intervene through their representatives in the procedure for amending a social network's terms and conditions, and the charters applicable to special regimes.

### Selection of a user representative

Account holders have the right to select user representatives on a regular basis through a public and transparent procedure. This could be a regional election by users, or a draw taking into account the different categories of users.

### Interoperability of data

Users should have the opportunity to choose the degree of interoperability (the ability of computer systems or software to exchange and make use of information) relating to their data, based on open standards.

### Rights over data

Users' data is the raw material behind the social networks' business model. "User-held data meets all the requirements of an 'asset' in property laws regardless of the fact that data could be deemed as being intangible".<sup>32</sup> Thus, users should have control over their data. This means being able to:<sup>33</sup>

- access data to review what information has been collected or inferred;
- erase data from a device and server;

<sup>32</sup> Jurcys, Paulius, et al. (2020).

<sup>33</sup> The World Wide Web Consortium (2022).

- transfer personal data to another platform;
- correct data to ensure that one's identity is properly reflected in a system;
- be protected in case of identity-related crime (identity theft and identity fraud);
- be free from automated decision-making, in order to exclude oneself from automated profiling;
- object, withdraw consent, and restrict use in case one changes one's decision.

---

## 2.2 OBLIGATIONS OF ACCOUNT HOLDERS

---

**Users should also commit to following the below obligations when interacting on social networks.**

### **Individuals liable for content**

Participants in the public debate are responsible for the content they disseminate or contribute to disseminating. In addition, given the effect of virality on social media, we recommend holding account holders liable for amplifying harmful content, even if they are not its author, except when the content discloses public interest information. The public interest quality of content should be determined by an independent judicial authority.

### **Identifying account holders**

As stated in the International Declaration on Information and Democracy, the accountability of all participants in the public debate is an essential principle that implies transparency about their identity. Under a principle of individual responsibility, each participant is responsible for the content they post on platforms. It is therefore essential to be able to identify the person behind the post.

Nevertheless, exceptions to the principle of transparency may be legitimate if they facilitate critical information pertaining to the public interest or if they contribute to the safety of participants. Indeed, anonymity can represent a guarantee allowing participants to express themselves.

We recommend tempering the principle of guaranteeing anonymity when necessary by allowing the possibility of revealing the identity of the person under specific and restrictive conditions.

Users have a right to be informed if a social network has initiated a procedure to reveal their identity, and can request assessment by international independent bodies such as the Human Rights Committee.

### **Mass-reporting**

Users must not engage in mass-reporting. They must adhere to the limits on the number of reports that can be submitted in a specific time period.

### **Impersonation of individuals, groups or organizations**

Account holders may not impersonate other individuals, groups or organizations to mislead others. However, parodies or fan pages may qualify for exemptions. In this case, these types of accounts must be explicitly presented as such.

### **Platform manipulation**

Users may not manipulate the platform to alter access to information or artificially amplify content. This includes spam or other activities that manipulate user experience.<sup>34</sup>

---

34 Tony Blair Institute (2021).

### Electoral and civic integrity

Users must not use social media networks with the intent of manipulating electoral or democratic processes.

### Misinformation

Users must not share misinformation that may result in imminent physical harm, or share false and misleading information in times of crisis, such as social armed conflicts, public health emergencies, and natural disasters.

---

## 2.3 SANCTIONS

---

Account holders must make sure they comply with a social network's terms of service. In case of non-compliance, users must be held accountable. Sanctions imposed by social networks must be made in accordance with human rights, i.e., the sanction must have a legal basis, pursue a legitimate aim, and be proportionate.

The criterion of proportionality – also called the necessary and proportionate test – requires that the choice of sanction be based in particular on the following criteria: nature and seriousness of the act, harm suffered by the victim, intention of the perpetrator of the act, risk of recidivism, and apology to the victim, among others.

Many legal provisions already cover certain sanctions attached to breaching terms and conditions, but do not target the means used to realize the activity itself on the social network. The account holder who has carried out an illegal act or breached the terms and conditions, regardless of its gravity, retains the means of committing an illegal act on the social network after being found responsible. In order to make account holders responsible and to prevent repeated breaches, we recommend the adoption of a wider framework of sanctions intended to prevent account holders from continuing to commit illegal activity by removing the possibility of their using their account.

Account holders have a right to appeal sanctions before an independent regulator or to the national judicial authority.

## SCALE OF SANCTIONS FOR USERS

### Scale of Sanctions According to Specific Behavior (cumulative and additional to criminal penalties)

A warning + indication	A limitation of the scope of diffusion	Removal of the Content	A limitation of the functionalities of the social network	A temporary suspension of the account < 1 year > 1 year > 5 years > 10 years	A suspension of access to platforms for up to 10 years
An isolated breach of terms and conditions		A repetition of breach of terms and conditions			A high frequency of breach of terms and conditions
					Death threat Hate speech Incitement to violence Cyber-bullying
Breach from Highly Influential Accounts			Breach of the obligation of being identifiable and transparent about sponsors		Repetitive breach of the obligation of being identifiable and transparent about sponsors
Non-compliance with advertising obligations			Repetitive non-compliance with advertising obligations		



### RECOMMENDATIONS TO STATES

Ensure that social networks recognize and protect the rights of account holders.

- > In cases of non-compliance, states can impose sanctions, as detailed in Chapter 1, Section 1.3.
- > Provide a mechanism for account holders to appeal social networks' moderation practices and sanctions.
- > Hold account holders liable for committing online crimes.
- > Provide a judicial procedure to determine reparations for victims. In addition, states should set up a national solidarity fund funded by social networks to compensate victims in case the perpetrator of the act is unable to do so for financial reasons.

# Chapter 3:

# Special Regimes

Deeply convinced that everyone should be equal before the law, we reaffirm the recognition of the principle of equality before the law guaranteed by Article 26 of the International Covenant on Civil and Political Rights (ICCPR).

However, a difference in treatment is not discriminatory if the measure at issue is based on a legitimate objective. Not only is the difference in treatment then allowed, but it is morally required. Indeed, some categories of account holders are in situations that require that they receive different treatment. The importance of the role of certain categories of account holders (e.g., those in favor of the protection of human rights or the right to information) requires that they benefit from additional procedural guarantees. Some account holders must bear greater obligations, because of their influence in the democratic debate and therefore their capacity to cause harm.

Consequently, **we recommend the establishment of special liability regimes for certain categories of users** because of the importance of their role in the public debate, the different obligations incumbent on them, or their strong capacity for influence and causing harm.

---

## 3.1 NEWS MEDIA AND JOURNALISTS

---

Journalism is facing an existential threat. News media and individual journalists' publication on platforms can be moderated or invisibilized, leading to censorship. These moderation decisions take place without consultation and therefore without adversarial debate.

It is necessary to provide a liability regime with additional guarantees to enable news media and journalists to fulfill their role in our democracies.

### ADDITIONAL RIGHTS

#### Specific status

News media should hold specific status. Social media platforms should recognize that journalistic media outlets have a specific social function, and that their accounts should benefit from some safeguards and protections, on condition that the media implement journalistic principles. News media should be identified as such by bodies independent of government, according to non-discretionary processes, without any political interference, using transparent self-regulatory standards and compliance mechanisms as much as possible.

#### Preserving visibility

Considering the editorial responsibility of news media, journalistic content should not be down-ranked, demoted, or removed. Moderation decisions taken against news media and journalistic content should not negatively affect the visibility of their posts or their monetization.

#### Third-party review

Moderation of journalistic content should only take place in the context of a procedure carried out by an independent authority, at the news media's own initiative, or at the request of any interested person, including through the platform. The procedure should be in the form of an expedited appeal.<sup>35</sup>

#### Information about changes to algorithms affecting financial stability

News media and journalists have the right to be informed of changes to the algorithms that could jeopardize their financial stability.

#### Fair compensation

Given that a platform's algorithms can impact the financial stability of media by reducing the visibility of their content, news media have the right to seek compensation for the damage suffered by changes to the algorithm, in order to ensure their financial viability.

---

35 Aspen Institute (2021).

## 3.2 MINORS

Platforms are taking an increasingly important place in most aspects of children's lives. It is therefore necessary to establish a specific liability regime that takes into account a user's age and ability.

**We must also recognize minors' rights to freedom of expression on social media**, in accordance with Article 13<sup>36</sup> of The International Convention on the Rights of the Child (1989). This right includes the freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, in oral, written, printed or artistic form, or by any other means of the child's choice.

### ADDITIONAL RIGHTS

#### Access to a legal regime

Minors have the right to access a legal regime based on the child's higher interest.

#### Access to child-friendly content

Minors have the right to access child-friendly, safe channels and content. They also have the right to safely seek advice and help regarding sexually explicit content generated by themselves or that they have received.

#### Access to age-appropriate procedures

Minors have the right to access age-appropriate procedures based on their best interest: for example, platforms can integrate trusted third parties to the subscription<sup>37</sup> of a minor so that they can confirm the identity or age of the user without revealing any identifying information. Moreover, platforms can be encouraged to study the feasibility of an age verification system<sup>38</sup> based on the operating system of a smartphone, tablet or computer, which can only be modified by the operator or by the parent upon presentation of proof of parentage.

## 3.3 INFLUENCERS

Platforms have allowed the emergence of a new profession: the influencer. This is an individual who, thanks to their exposure on the internet, has an influence over users who follow them and over their purchasing decisions. Influencers differ from a simple account given their high visibility and the fact that they carry out a paid activity. The economic benefits they receive from this exposure as well as its strong ability to influence the choices of users justify the establishment of a specific liability regime.

We recommend that social networks recognize influencers as users who are receiving money or any kind of reward in return for the promotion of a product or service through their influence on a platform.

36 "The child shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of the child's choice."

37 CNIL (2022).

38 Institut Montaigne (2021).

## ADDITIONAL RIGHTS

### Demotion and removal of content

Given that a platform's moderation decision can impact the financial position of influencers by reducing the visibility of their content, influencers have the right to be informed when a moderation decision or a change to the algorithms can have a financial impact on them.

## ADDITIONAL OBLIGATIONS

### Code of conduct

Influencers must adhere to a code of conduct imposed by the platform, framing their activity according to clear principles.

### Sponsored content

All content that has been sponsored and posted by an influencer account, should be presented as such to all users in the social media network.

### Disclosures regarding advertising agreements

Influencers should disclose all advertising agreements, as well as the amount of money received and the targeting strategy on political advertising. These disclosures should include information such as:

- the location and source of payments;
- the size of the target audience, as well as the number of views the advertisement receives, together with user engagement beyond viewing the advertisement;
- selection criteria for targeting recipients (while protecting privacy) as is communicated to advertisers, including the data source, inferred profile, lookalike audiences, custom audiences, and A/B testing practices;
- the revenues from targeted advertising.

## SANCTIONS

### Scope of sanctions

Sanctions must reflect an influencer's level of influence. The determination of the applicable sanction should take into account the fact that those account holders act as professionals and must therefore bear an obligation reinforced by their influence.

## 3.4 HIGHLY INFLUENTIAL ACCOUNTS

Highly visible accounts, although not necessarily operating on platforms to make a profit, have a strong influence over users because of the size of their community, which leads to a high visibility of their content. This visibility necessarily increases the risk of causing greater harm, requiring stronger oversight to prevent or stop breaches more quickly.

An account is considered highly influential when it meets one or more of the following conditions: it has **50,000 followers or more; is an account that belongs to someone who performs a representative function (politicians, public authority officials);** or some other public figure who has been certified by the platforms.

## ADDITIONAL RIGHTS

### Accelerated appeal process

Highly influential accounts should have the right to access an accelerated appeal process to decide moderation decisions and sanctions.

## ADDITIONAL OBLIGATIONS

### Code of conduct

Public figures and politicians must adhere to a code of conduct imposed by the platform, framing their activity according to clear principles, such as being identifiable and transparent about their political or company affiliation.

### Infringement on others' rights

It should be recognized that public figures have a strengthened obligation not to infringe on the rights of others, including the right to information.<sup>39</sup>

### Disclosures regarding advertising agreements

Highly influential accounts should disclose all advertising agreements, as well as the amount of money received and the targeting strategy on political advertising.<sup>40</sup> These disclosures should include information such as:

- the location and source of payments;
- the size of the target audience, as well as the number of views the advertisement receives, together with user engagement beyond viewing the advertisement;
- selection criteria for targeting recipients (while protecting privacy) as is communicated to advertisers, including the data source, inferred profile, lookalike audiences, custom audiences, and A/B testing practices;
- the revenues from targeted advertising.

## SANCTIONS

### Scope of sanctions

Sanctions must reflect a highly influential account's level of influence and capacity to cause harm. The determination of the applicable sanction should take into account their strong capacity for influence and therefore harm. It should take into account the context and the visibility of the content.

---

## 3.5 GROUP ADMINISTRATORS

---

The management of certain content on platforms is carried out by a group administrator. Their role is to supervise, manage and administer a group or account on a social network, giving them significant moderation power. Consequently, group administrators may have the capacity to act to stop the dissemination of illegal or harmful content in their group. Thus, group administrators should bear specific obligations and receive additional rights.

---

<sup>39</sup> Brennan Center for Justice (2021).

<sup>40</sup> Reisach, Ulrike (2021).

## ADDITIONAL RIGHTS

### A members' charter

Group administrators have the right to adopt and enforce a charter on their members, which, provided that it complies with human rights and the platform policy, should be able to include specificities justified by the purpose of the group.

### Active moderation of content

Group administrators have the right to actively participate in moderation of content within their group.

## ADDITIONAL OBLIGATIONS

### Code of conduct

Group administrators must follow a code of conduct imposed by the platform, framing their activity according to clear principles.

### Cooperation with the platform and authorities

Group administrators should be obliged, when informed of an illegal act posted on their group, to cooperate with the platform and the competent authorities to ensure that the individual puts an end to their activity and is prosecuted.



### RECOMMENDATIONS TO STATES

- > Ensure that social networks recognize special regimes and establish procedures accordingly.
- > In cases of non-compliance, states can impose sanctions, as detailed in Chapter 3, Section 1.3.
- > Adopt a common definition of influencers by taking into account their professional activities on social networks.

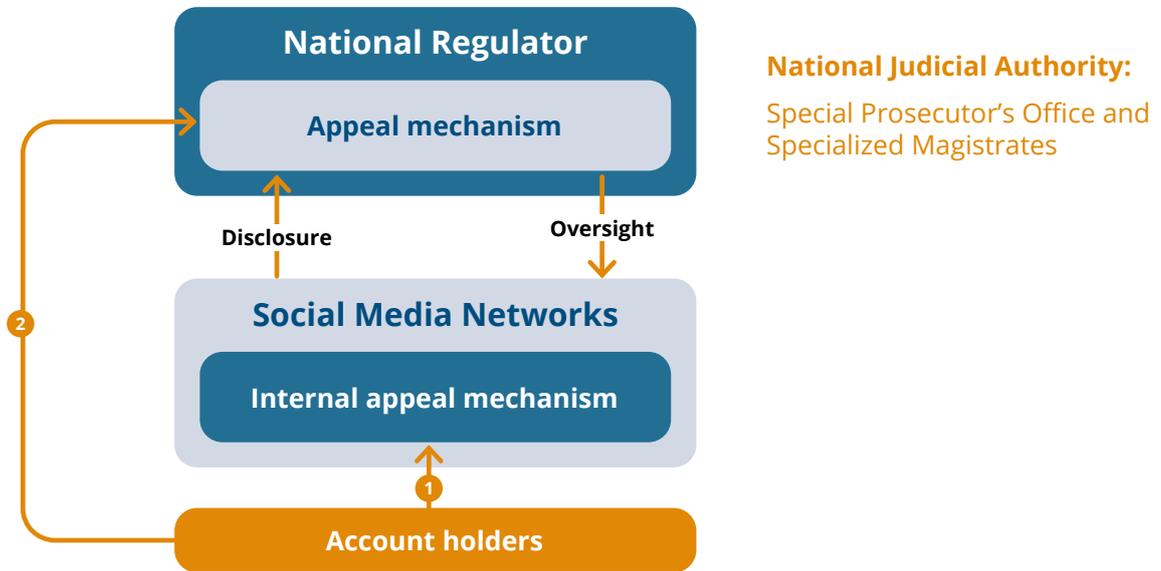
# Chapter 4:

# Governance

The International Declaration on Information and Democracy states that the accountability of all participants in public debate is an essential principle. In order to implement the liability regimes set out in the previous chapter, governance bodies and procedures must be put in place.

Social networks are by nature a transnational activity requiring cooperation between the different states and participants to effectively mitigate illegal behavior online and preserve democracy. It is therefore essential to strengthen the national governance through the establishment of international bodies.

## 4.1 NATIONAL LEVEL



The first level of action for the implementation of accountability regimes should be at the national level. It would encompass an independent regulator, a mechanism for reporting illegal behavior or content, access to an independent judicial system and a specialized prosecutor's office and magistrates.

### 4.1.1 INDEPENDENT REGULATOR

We recommend the establishment of an independent regulator, or the strengthening of an existing independent administrative body, to monitor, promote and protect the rights of users on platforms and to resolve conflicts between users and the platform.

#### MANDATE:

We recommend that this authority should have four competences:

#### 1. Platform Oversight

**We recommend that this authority ensure enforcement and compliance by platforms,** unless specific tasks have been entrusted to other competent authorities.

We recommend that this oversight be exercised through the following different roles.

The relevant authority should:

- receive disclosure reports from platforms on their practice and evaluate them;
- have certain investigative powers, such as the power to require information from suppliers, conduct on-site inspections, and ask employees to explain their decisions;
- be competent to monitor the general terms and conditions (an ex-ante check) to ensure that platforms comply with all applicable legislation;
- have the power to impose sanctions on platforms;
- be competent to clarify any issue about identification of specific accounts, including highly influential

accounts, public figures, or influencers (see Chapter 3);

- be able to issue recommendations, in order to help platforms adapt their practices;
- review algorithmic impact assessments and supervise frequent stress tests.

The decisions of this authority shall be binding on all matters concerning access to information, including the resolution of related disputes.

These powers are administrative in nature and do not seek to replace those of courts.

The authority should have the power to sanction platforms that do not exercise best efforts in accordance with high industry standards of professional diligence **to avoid the availability of illegal content, harmful disinformation, and unauthorized works on their services.**

#### **The authority should hold platforms responsible for:**

- not acting expeditiously upon obtaining such knowledge or awareness, to remove or to disable access to illegal online content, whether it is hate speech or an illegal product;
- not cooperating with the national authority;
- repetitive abusive deletions of content / multiple errors of the AI-based moderation;
- absence of dispute mechanisms that enable users to appeal a decision;
- violation of transparency obligations owed to national or international authorities and to users and vetted researchers;
- denying users the possibility to choose an algorithm that is not based on a behavioral profile;
- development, use and application of artificial intelligence, algorithms, and other similar technologies manifestly incompatible with international human rights law and standards;
- use of “dark patterns”, which design choices that mislead or steer people into decisions they may not otherwise have made;
- not proceeding to annual audits of how their algorithms affect democracy, human rights, and the physical and mental health of minors and other users;
- deceptive commercial practices if their public communication is not in line with their moderation practices, including in the event of a change in rules without warning;
- profiling of minors, and the use of sensitive data pertaining to sexual orientation, religion, and ethnicity for advertisement purposes.

## **2. Advisory role**

**We recommend that this authority gives advice and participates in the development of public policies and regulations at the national level.** We encourage the creation of a multi-stakeholder body within the administrative authority to develop common principles, rules, procedures based on international regulatory recommendations and standards, and decision-making programs to define the use and evolution of the internet.

- **This national body should have a consultative competence.**
- **It should be consulted on each legislation project that has an impact on platforms.**
- It should be composed of experts with recognized competence in the field of moderation and platforms.
- Its members will have to represent the various parties interested in this issue (civil society, media, NGOs, etc.).

## **3. Support for users of digital services**

**We recommend that this authority ensure the role of intermediary with the platforms by:**

- connecting users and platforms;
- helping and guiding victims;
- handling appeals by account holders. We recommend that these decisions be binding on the platform.

## 4. Research cluster

Platforms hold a monopolistic control over research access to their services. As such, **we recommend the creation of a research cluster** within the administrative authority. It would bring together researchers and digital specialists to advance knowledge of algorithmic content moderation systems, develop new ones, audit those of the main platforms, and link platforms and researchers to improve access to platform data in anonymized format.



### IMPLEMENTATION BY STATES

- > **The members of this public regulatory body must be independent.**
- > **Their nomination process must be open, transparent, and involve the participation of relevant stakeholders.**
- > **The members selected should be individuals of high moral character and recognized competence in the fields of technology, including individuals with legal experience.**
- > **All interested parties including civil society must be fairly represented.**
- > **A specific taxation on large digital platforms should be used to finance this authority.**

## 4.1.2 MECHANISM FOR REPORTING ILLEGAL BEHAVIOR OR CONTENT

**We recommend the establishment of a mechanism for reporting illegal behavior or content online.**

A great deal of illegal content posted on platforms is moderated without national authorities being informed, or is only moderated if a victim initiates proceedings. The removal of content or its invisibilization may indeed be necessary to put an end to the harm suffered by the victim. However, it may then be difficult for the authorities to launch an investigation in the absence of knowledge or evidence of the disputed issue. Similarly, victims face the difficulty of proving the content, especially if it has been removed.



### IMPLEMENTATION BY STATES

- > **We recommend that states identify a centralized authority competent to receive and evaluate content deemed illegal by a platform when necessary, for example in the case of user appeal. This can take the form of an existing authority such as a specialized prosecutor's office, an independent agency or a new platform for reporting illegal behavior or content on the internet.**
- > **If an independent agency is entrusted with this role, we recommend that a cooperation between this agency and the state prosecutor is established, to facilitate the launching of investigation.**
- > **We recommend simplifying the degree of proof required of victims reporting illegal content which has already been removed. This could be achieved by a judicial authority setting up a free reporting procedure.**

### 4.1.3 ACCESS TO INDEPENDENT JUDICIAL SYSTEM

**We recommend that access to independent jurisdiction be provided for civil proceedings.**

Jurisdiction *ratione loci* (i.e., jurisdiction subject to territorial boundaries) confines litigation to the place where the platform has its registered office, and this raises difficulties in determining the jurisdiction of national courts. We recommend that this criterion, which is not relevant to the international activity of the platforms, be disregarded. This approach is also consistent with the case-law of national courts.

Due to the particularity of this litigation requiring specific knowledge, and significant resources, it is necessary that judges involved in this litigation are specialized.

Due to the importance of the irreparable damage that can be caused by the dissemination of content, it is vital to enable judges to promptly order hosts to remove content, or the Internet Service Provider (ISP) to block access to the content.

This expedited removal of content procedure could be opened either by the Public Prosecutor's Office or more largely by any person having an interest in it.

Anonymity is a considerable obstacle to carrying out procedures. Under certain circumstances, specific procedures should permit obtaining the identity of the person in a short time.



#### IMPLEMENTATION BY STATES

- > **We recommend that states recognize the right of users to appeal against a decision taken by a platform against them before a national civil court.**
- > **We recommend that states exercise their authority on the basis of a criterion of *ratione personae*, by retaining the jurisdiction of the national court of the place in which the victim resides, or of the user who posted the illegal content.**
- > **We recommend that, when possible, states make use of their universal jurisdiction to participate in the fight against impunity for the most serious crimes.**
- > **National judges engaged in litigation involving social networks should have particular qualities such as impartiality and independence.**
- > **States should grant sufficient human and financial resources to judicial institutions.**
- > **We recommend the establishment of an independent civil court specialized in litigation involving platforms. Its composition should include representation of civil society.**
- > **We recommend the establishment of summary proceedings (ie., a civil or criminal proceeding conducted without formalities such as pleadings) for the speedy disposition of a matter.**
- > **States should set up a procedure on request to identify the author of disputed content.**

## 4.1.4 SPECIALIZED PROSECUTOR'S OFFICE AND MAGISTRATES

Given the detailed nature of the litigation in question, we recommend the creation of specialized magistrates and specialized prosecutors.

## 4.2 INTERNATIONAL LEVEL

### TRANSNATIONAL BODY

#### REGULATION

1. Set standards for platforms.
2. Identify exceptional circumstances.
3. Systemic analysis of the practices of platforms in order to contribute to the improvement of regulation.
4. Participate in the oversight of platforms.
5. Oversee states' obligation to respect freedom of expression.
6. Expert body to provide advice at the request of platforms.
7. Assist in the training of moderators .

#### DISPUTE SETTLEMENT

Provide platforms with a more flexible dispute resolution mechanism than domestic court proceedings to determine the application of the law.

#### INVESTIGATIVE ASSISTANCE

1. Collect, consolidate and preserve information and evidence of human rights violations and abuses.
2. Analyze the collected evidence in order to facilitate fair and independent criminal proceedings.
3. Share information and evidence collected, and analytical work produced with national, regional and international courts.

The global space of information and communication is a common good of humanity that must be protected and regulated at the international level, taking into account cultural diversity but also the common principles accepted and recognized by the various participants.

We recommend establishing a transnational body, given that national mechanisms may not meet the requirements imposed by international standards and should be monitored to ensure respect for the rights of account holders. This transnational body would have three separate mandates: international regulation, a settlement mechanism, and investigative assistance.

In addition, the law applicable to platforms stemming from international human rights law is not sufficiently precise as it stands, and should be refined. Finally, platforms need to clearly know the applicable law in order to carry out a fair moderation of illegal content.

## 4.2.1 INTERNATIONAL REGULATION

The transnational body should be composed of representatives of all participants including representatives of platforms, civil society, states, and interested international organizations, in order to promote regulation while ensuring coordination between the different stakeholders.

Indeed, the establishment of accountability regimes requires extensive work that our working group has not been able to carry out. This report therefore recommends that this mission be carried out by a standing body. The latter will be able to determine more specifically the criteria dictating the application of sanctions in the light of the circumstances in particular cases, in order to guide platforms and national authorities in the regulation of social networks. Its permanent nature will allow it to adapt regulation in the light of changing situations.

This transnational body will provide support to platforms to implement national and international regulations.

This body could be created within an existing organization. The Forum on Information and Democracy already has the necessary structure to exercise the competences required to promote and control the regulation of platforms in accordance with international standards.

### Setting up a transnational body

#### MANDATE:

#### 1. Advising and guiding platforms

- Specify the main principles that platforms must comply with.
- Establish the means of implementation of such principles through the development of standards and guidelines.
- Guide platforms' response in exceptional situations such as war.
- Provide advice at the request of platforms.

#### 2. Overseeing platforms' practices

- Evaluate the decisions made by platforms through peer review and/or carry out a specific investigation based on a complaint.
- Publish the results of these evaluations in an effort of transparency allowing, on the one hand, platforms to increase their legitimacy, and on the other, users to make an informed decision on the social network they want to use.

#### 3. Overseeing States' obligation to respect freedom of expression

- Issue, on a regular basis, a ranking of states based on their respect for human rights and the fundamental freedoms of journalists and human rights defenders. This ranking would help the platforms to determine if they can reveal the identity of an account holder.
- Review the activity of the national regulatory authorities, in order to verify that they benefit from the safeguards essential to carry out their functions in an independent, impartial and transparent manner.

## GOVERNANCE

The transnational body should include representatives of all stakeholders in the information and communication space.

It should be composed of the following organs:

- A **plenary body** bringing together representatives of all parties concerned.
- An **administrative body** composed of people with expertise in the field, respecting an equitable geographical and gender representation.
- An **expert body** that should be mobilized for the oversight of platforms and consulted during the procedures of designing standards and guidelines.
  - > Members of the expert body should be elected by the plenary body for their specialized competence and respect the principle of independence.

## 4.2.2 INTERNATIONAL DISPUTE SETTLEMENT

Famous cases have highlighted the importance of providing platforms with a more flexible dispute resolution mechanism than domestic court proceedings to determine the application of the law. We recommend the creation of an international dispute settlement mechanism based on arbitration principles.



### IMPLEMENTATION BY STATES

- > **We recommend the creation of an arbitration<sup>41</sup> system within the transnational regulatory body.**
- > **It should provide:**
  - **a list of arbitrators: this list should be drawn up taking competences into account, and could be approved annually by the plenary body.**
  - **a procedural guide: the organization should offer a procedural guide that the parties can then adapt.**
- > **The parties should undertake that the decision will be binding.**

### Investigative assistance

Information posted on platforms can be used as evidence of either violation of human rights or crimes. However, the moderation obligation incumbent on platforms can lead to the loss of this evidence.

Even if platforms agree to conclude agreements with international investigative mechanisms and to transmit information posted on their services, these agreements are ad hoc and limited in time. They do not prevent the loss of evidence in the absence of international mechanisms.

The investigative assistance capacity of the international body should serve as a fact-finding body to establish facts and help decide the cases, or also to assist other bodies, domestic or international. This should be done in connection with the arbitration process and other existing or to be created bodies in charge of gathering evidence on crimes.

<sup>41</sup> Strowel, Alain & Laura Somaini (2021).



## IMPLEMENTATION BY STATES

- > **We recommend the creation of an investigative assistance body within the transnational body.**
- > **This body should be composed of competent people: legal, researchers, and analysts, among others, recommended by an independent head of the transnational body and elected by the plenary body.**
- > **This body should be given the mandate to:**
  - **collect, consolidate and preserve information and evidence of human rights violations and abuses;**
  - **analyze the collected evidence and prepare files in order to facilitate and expedite fair and independent criminal proceedings;**
  - **share information and evidence collected and analytical work produced with national, regional and international courts.**

# APPENDIX 1

## INCOMING ACCOUNTABILITY REGIME IN EUROPE: THE DIGITAL SERVICES ACT

This report highlights the need for accountability regimes for social media networks and their users and proposes a path forward. The European Union is moving in this direction with the introduction of the Digital Services Act (DSA).<sup>42</sup> The DSA is an unprecedented regulatory framework that will set standards for the accountability of online platforms regarding illegal and harmful content and will provide better protection for internet users and their fundamental rights. The DSA is surfacing as one of the regulatory approaches in this field, and could serve as a blueprint for other jurisdictions as countries move forward in implementing democratic safeguards to the information and communication space.

The DSA will address the following issues, which are also covered by this report:

### For users:

- Create easy and clear ways to report illegal content.
- Users are informed about, and can contest removal of content by platforms.
- Users will have access to dispute resolution mechanisms in their country.
- New protections for minors.
- Access to platforms' data for researchers to understand risks for society and fundamental rights.
- Users' right to opt-out from content recommendations based on profiling.

### For platforms:

- Transparent terms and conditions for platforms.
- Fast crisis response mechanisms with additional risk-management measures for public health and security crises.
- Independent auditing of their risk management, including for their algorithmic systems.
- Transparency of the rules for content moderation.
- Clear information on why content is recommended to users.

---

<sup>42</sup> European Parliament (2022).

## ACKNOWLEDGMENTS

The Forum would like to warmly thank the Steering Committee members of this working group, especially its co-chairs, and the more than 50 experts we interviewed, for their time, insightful perspective and support.

- Camille Alloing, professor of public relations, University of Québec
- Enrique Armijo, professor of Law, Elon University School of Law and a Faculty Affiliate of the Yale Law School Information Society Project
- Farzaneh Badiei, founder, Digital Medusa
- Romain Badouard, lecturer, Paris 1
- Ian Barber, senior legal officer, Global Partners Digital
- Dominique Cardon, director, Médialab Sciences Po
- Anupam Chander, Scott K. Ginsburg Professor of Law and Technology, Georgetown University Law Center
- Paul Coppin, head of legal desk, Reporters Without Borders (RSF)
- Alexandre Coutant, professor of social and public communication, University of Québec
- Vera Eidelman, staff attorney, ACLU Speech, Privacy, and Technology Project
- Olivier de Frouville, director, Paris Center for Human Rights (CRDH)
- Anna D. Gibson, PhD candidate, Department of Communication, Stanford University
- Eric Goldman, professor of law, Santa Clara University School of Law
- Vincent F. Hendricks, professor, director, Center for Information and Bubble studies, University of Copenhagen
- Matias Jackson, legal researcher and consultant on technology and internet law, University of the Republic of Uruguay
- David Kaye, former special rapporteur on freedom of expression and opinion (2014-2020)
- Julien Kahn, lawyer, JK Avocats
- Théophile Lenoir, PhD, Leeds University
- Richard Malka, lawyer, Malka & Associés
- Barthélémy Michalon, PhD, Sciences Po
- Charles Mirailié, legal advisor, National Consultative Commission for Human Rights
- Gabriel Nicholas, research fellow, Center for Democracy & Technology
- Rasmus Nielsen, director, Reuters Institute for the Study of Journalism
- Javier Pallero, policy director, Access Now
- Courtney Radsch, advocacy director, Committee to Protect Journalists
- Aurore Raoux, public affairs officer, News Media Europe
- Emilie Rannou, senior manager in strategy & data science, Ekimetrics
- Alain Strowel, professor, University Saint-Louis and UCLouvain
- Nicolas Suzor, professor, Law School at Queensland University of Technology
- Hélène Tigroudja, professor of international public law, University Aix Marseille
- Erik Tuchtfeld, research fellow, Max Planck Institute for Comparative Public Law and International Law
- Serena Villata, tenured researcher, CNRS Chair in Artificial Intelligence
- Wout van Wijk, director, News Media Europe
- Richard Wingfield, head of legal, Global Partners Digital
- Jillian York, director for international freedom of expression, Electronic Frontier Foundation
- Célia Zolynski, member, National Pilot Committee for Digital Ethics

To strengthen the participation of experts from different regions in the working group, the Forum formalized a partnership with six local organizations in the MENA region and Western Africa:

- Maharat Foundation, a non-governmental organization for freedom of expression and media development based in Lebanon
- The Working Group for Journalism and Democracy, an organization promoting journalism and digital media development in Tunisia
- The Association for the Fight against Racism, Ethnocentrism and Regionalism (ALCRER), a non-governmental organization defending human rights, democracy and good governance in Benin
- The Network of Online Press Professionals of Côte d'Ivoire (REPPRELCI), the first professional organization of the digital press in Ivory Coast
- The Union of Information and Communication Professionals of Senegal (SYNPICS), a professional organization in charge of the media sector in Senegal
- New Human Rights (NDH), an association defending human rights and press freedom in Burkina Faso

Thanks to their support, 53 experts were mobilized and participated in workshops to develop the recommendations presented in this report:

#### **Lebanon:**

- Layal Bahnam, Program Manager, Maharat foundation
- Dr Ali Rammal, Media Professor, Lebanese University
- Dr Georges Sadaka, Media Professor and Former Dean of the Media Faculty, Lebanese University
- Abed Kataya, Digital Content Manager, SMEX for Digital Rights.
- Dr Maria Abou Zeid, Associate Professor and Chairperson, Media Studies Department at Notre Dame University-Louaize (NDU)
- Zeina Bou Harb, Head of International Cooperation, OGERO Telecom, currently heading the General Secretary of the Lebanese Internet Governance Forum.
- Abir Chebaro, gender expert, former advisor on women's affairs to the prime minister
- Layal Jebran, tech expert and entrepreneur, CEO and co-founder at Moubarmij, former executive board member at Internet Society Lebanon
- Charbel Chbeir, lawyer, member of the Beirut Bar Association and rapporteur for the ICT (Center for Informatics and Technology of the Beirut Bar Association)
- Layal Sakr, lawyer and founder of the NGO SEEDS for Legal Initiatives.
- Tony Mikhael, lawyer and legal expert of the NGO Maharat Foundation

#### **Tunisia:**

- Neji Bghouri, head of the UNESCO project office in Tunis and former president of the National Union of Tunisian Journalism
- Dr. Sadok Hammami, professor at the Institute of Press and Information Sciences and member of the press council
- Dr. Nouha Belaid, doctor in information and communication Sciences, professor at the Central University of Tunis.
- Karim Wannas, journalist at Mosâïque FM, responsible for ethics at the National Union of Tunisian Journalists
- Rabeab Aloui, editor-in-chief of the electronic media specialized in fact-checking, BN Check
- Lasaad Dahech, expert to the ASBU (Arab States Broadcasting Union) and former CEO of the National TV establishment
- Réal Barnabé, media expert for the European Commission, former Chief of Radio Canada
- Tawfiq Omrane, journalist and caricaturist

**Benin:**

- Gervais Loko, expert in governance and program manager at ALCRER
- Hugues Tossa, cyber-lawyer and project manager at Changement Social Benin (NGO)
- Renaud Sédjro Akpahounka, expert designer of CivicTech Solutions and ICT manager at Social Watch Benin (NGO network)
- Maurice Thantan, digital project manager and former president of the Benin Bloggers Association
- Donald Azokli Assogba, SEO (Search Engine Optimization) expert
- Hadjara Idris, digital expert and CEO, Mara Academy
- Constant Langnito, consultant expert Web et Mobile
- Joël Arsène Noumonvi, president of the Benin Bloggers Association
- Hervé Mekoun, sociologist and university professor

**Ivory Coast:**

- Lassina Serme, president of The Network of Online Press Professionals of Côte d'Ivoire (REPPRELICI)
- Kébé Mamady, editor-in-chief of the platform against fake news Ivoirecheck
- Marie Flore Begou, coordinator of the NGO Citizen Committed to Leadership and Democracy in Côte d'Ivoire (CELDICI),
- Marlyse Konan, president of the Association of Women Journalists of Côte d'Ivoire (AFJCI)
- Dr Karim Wally, teacher-researcher at the University Félix Houphouët-Boigny of Abidjan
- Aboubacar Cidiq Sawadogo, president of the Citizen Initiative for Social Media Education in Côte d'Ivoire (ICEM-CI).
- Evelyne Déba, journalist
- Wally Karim, editorial director of Nord-Sud Quotidien
- Elisée Bolougbeu, director of publication at Destinationci

**Senegal:**

- Bamba Kassé, journalist, President of Synpics
- Adama sow CDP, journalist at the Personal Data Protection Commission of Senegal (CDP)
- Moustapha Fall, lawyer specialized in public law, drafter of the Social Networks Regulation Bill
- Mamadou Thior, president of Cored (Council for the Regulation of Ethics and Professional Conduct)
- Ibrahima Lissa Faye, president of APPEL (Association for the Promotion of the Online Press)
- Emmanuel Maurice Diokh, techno-pedagogue, legal trainer, president of Internet Sans Frontières
- Aminatou Diop, journalist member of CORED (Council for the Observance of the Rules of Ethics and Professional Conduct in the Media in Senegal)
- Jean Louis Correa, associate professor of law at the Virtual University of Senegal

**Burkina Faso:**

- Idrissa Birba, journalist, chief editor of the newspaper Le Quotidien, president of NDH-Burkina
- Boukari Bogré, teacher and journalist
- Frédéric Ilboudo, journalist-communicator at the NGO Hunger Free World (HFW)
- Daouda Kinda, journalist, responsible for the online media, Le Monde Rural
- Madi Bazié, journalist, responsible for the online media, Ouest-info
- Élie Kaboré, journalist at the newspaper L'économiste du Faso
- Philippe Bationo, journalist, responsible for the online media Le Talon
- Zalle Ibrahim, journalist

# BIBLIOGRAPHY

- Access Now (2020). 26 Recommendations on Content Governance. Accessed via: <https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>
- ACLU (2021). The Problem With Censoring Political Speech Online – Including Trump’s. Accessed via: <https://www.aclu.org/news/free-speech/the-problem-with-censoring-political-speech-online-including-trumps/>
- American bar association (2018). The Ongoing Challenge to Define Free Speech. Accessed via: [https://www.americanbar.org/groups/crsj/publications/human\\_rights\\_magazine\\_home/the-ongoing-challenge-to-define-free-speech/](https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/)
- Article 19 (2021). At a glance: Does the EU Digital Services Act protect freedom of expression? Accessed via: <https://www.article19.org/resources/does-the-digital-services-act-protect-freedom-of-expression/>
- Aspen Institute (2021). Commission on Information Disorder. Accessed via: [https://www.aspeninstitute.org/wp-content/uploads/2021/11/Aspen-Institute\\_Commission-on-Information-Disorder\\_Final-Report.pdf](https://www.aspeninstitute.org/wp-content/uploads/2021/11/Aspen-Institute_Commission-on-Information-Disorder_Final-Report.pdf)
- Atlantic Council (2021). Free speech and online content: What can the US learn from Europe? Accessed via: <https://www.atlanticcouncil.org/blogs/new-atlanticist/free-speech-and-online-content-what-can-the-us-learn-from-europe/>
- Badouard, Romain (2021). Les enjeux de la modération des contenus sur le web. Accessed via: <https://la-rem.eu/2021/11/les-enjeux-de-la-moderation-des-contenus-sur-le-web/>
- Banko, Michele, Brendon MacKeen & Laurie Ray (2020). A Unified Typology of Harmful Content. Accessed via: <https://aclanthology.org/2020.alw-1.16.pdf>
- Bayer, J.A (2019). The recommended responsibilities and duties of social media platform companies. Accessed via: [https://www.researchgate.net/publication/344785231\\_The\\_recommended\\_responsibilities\\_and\\_duties\\_of\\_social\\_media\\_platform\\_companies](https://www.researchgate.net/publication/344785231_The_recommended_responsibilities_and_duties_of_social_media_platform_companies)
- Berinsky, A. J (2015). Rumors and Health Care Reform: Experiments in Political Misinformation, Cambridge University Press, Vol. 47, pp. 241 - 262. Accessed via: <https://www.cambridge.org/core/journals/british-journal-of-political-science/article/rumors-and-health-care-reform-experiments-in-political-misinformation/8B88568CD057242D2D97649300215CF2>
- Boyd, Danah M. & Nicole B. Ellison (2007). Social Network Sites: Definition, History, and Scholarship, Journal of Computer-Mediated Communication, Vol. 13, pp. 210–230. Accessed via: <https://academic.oup.com/jcmc/article/13/1/210/4583062>
- Brennan Center for Justice (2021). Double Standards in Social Media Content Moderation. Accessed via: [https://www.brennancenter.org/sites/default/files/2021-08/Double\\_Standards\\_Content\\_Moderation.pdf](https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf)
- Buiten, Miriam, (2021). The Digital Services Act: From Intermediary Liability to Platform Regulation, SSRN. Accessed via: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3876328](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3876328)
- Burgorgue-Larsen, Laurence (2008). Les nouvelles tendances dans la jurisprudence de la Cour interaméricaine des droits de l’homme, Cursos de Derecho Internacional y Relaciones Internacionales de Vitoria Gasteiz, Universidad del País Vasco, pp. 149-180. Accessed via: <https://hal.archives-ouvertes.fr/hal-01744225/document>
- Business for Social Responsibility (2021). A Human Rights-Based Approach to Content Governance. Accessed via: [https://www.bsr.org/reports/A\\_Human\\_Rights-Based\\_Approach\\_to\\_Content\\_Governance.pdf](https://www.bsr.org/reports/A_Human_Rights-Based_Approach_to_Content_Governance.pdf)
- Cassella, Sarah, (2018). Rapport introductif, Le standard de due diligence et la responsabilité internationale. Accessed via: <https://www.sfdi.org/publications/le-standard-de-due-diligence-et-la-responsabilite-internationale-journee-detude-franco-italienne-du-mans/>
- CIGI (2021). How “Free Speech” Kills Internet Regulation Debates: Part Two. Accessed via: <https://www.cigionline.org/articles/how-free-speech-kills-internet-regulation-debates/>
- CNIL (2019). La forme des choix, Cahier IP innovation & Prospective, Num.06. Accessed via: [https://www.cnil.fr/sites/default/files/atoms/files/cnil\\_cahiers\\_ip6.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil_cahiers_ip6.pdf)
- CNIL (2022). Contrôle de l’âge sur les sites web: la CNIL invite à développer des solutions plus efficaces et respectueuses de la vie privée. Accessed via: <https://www.cnil.fr/fr/contrrole-de-lage-sur-les-sites-web-la-cnil-invite-developper-des-solutions-plus-efficaces-et>
- Common, MacKenzie, (2019). The Importance of Appeals Systems on Social Media Platforms, LSE Law - Policy Briefing Paper, Vol. 40, pp. 1-5. Accessed via: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3462770](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3462770)
- Congressional Research Service (2019). Free Speech and the Regulation of Social Media Content. Accessed via: <https://sgp.fas.org/crs/misc/R45650.pdf>
- Congressional Research Service (2021). Social Media: Misinformation and Content Moderation Issues for Congress. Accessed via: <https://crsreports.congress.gov/product/pdf/R/R46662>
- Conseil de l’Europe (2014). La prééminence du droit sur l’internet et dans le monde numérique en général. Accessed via: [https://rm.coe.int/la-preeminence-du-droit-sur-l-internet-et-dans-le-monde-numerique-en-g/16808b3e6f\\_juin\\_2021](https://rm.coe.int/la-preeminence-du-droit-sur-l-internet-et-dans-le-monde-numerique-en-g/16808b3e6f_juin_2021)
- Council of Europe (2020). Prioritization uncovered: the Discoverability of Public Interest Content Online <https://rm.coe.int/publication-content-prioritisation-report/1680a07a57>
- Council of Europe (2021). Content Moderation: Best Practices towards Effective Legal and Procedural Frameworks for Self-regulatory and Co-regulatory Mechanisms of Content Moderation. Accessed via: <https://edoc.coe.int/fr/internet/10198-content-moderation-guide-ance-note.html>
- D’Ambrosio, Luca (2020). Le devoir de vigilance: une innovation juridique entre continuité et ruptures, Droit et Société, Vol. 106, pp. 633-647. Accessed via: <https://www.cairn.info/revue-droit-et-societe-2020-3-page-633.htm>

- Data & Society (2018). Content or context moderation: Artisanal, Community-Reliant, and Industrial Approaches. Accessed via: [https://datasociety.net/wp-content/uploads/2018/11/DS\\_Content\\_or\\_Context\\_Moderation.pdf](https://datasociety.net/wp-content/uploads/2018/11/DS_Content_or_Context_Moderation.pdf)
- Desmaris, Sacha, Pierre Dubreuil & Benoît Loutrel (2019). Créer un cadre français de responsabilisation des réseaux sociaux: agir en France avec une ambition européenne. Accessed via: <https://la-rem.eu/2021/11/les-enjeux-de-la-moderation-des-contenus-sur-le-web/>
- Douek, Evelyn (2020). The Rise of Content Cartels. Accessed via: <https://knightcolumbia.org/content/the-rise-of-content-cartels>
- Douek, Evelyn (2021). Governing online speech: from “post-as-trump” to proportionality and probability, Columbia Law Review, Vol. 121, pp. 759-834. Accessed via: <https://columbialawreview.org/content/governing-online-speech-from-posts-as-trumps-to-proportionality-and-probability/>
- Dugoin-Clément, Christine, (2021). Donald Trump banni des grands réseaux sociaux: et maintenant ? Accessed via: <https://theconversation.com/donald-trump-banni-des-grands-reseaux-sociaux-et-maintenant-152950>
- ECFR (2020). Regulation and accountability: How to save the internet. Accessed via: [https://ecfr.eu/article/commentary\\_regulation\\_and\\_accountability\\_how\\_to\\_save\\_the\\_internet/](https://ecfr.eu/article/commentary_regulation_and_accountability_how_to_save_the_internet/)
- Electronic Frontier Foundation (2021). Beyond Platforms: Private Censorship, Parler, and the Stack. Accessed via: <https://www.eff.org/deeplinks/2021/01/beyond-platforms-private-censorship-parler-and-stack>
- Electronic Frontier Foundation (EFF). How To Appeal. Accessed via: <https://www.onlinecensorship.org/en/pages/1214ccb8-e1d5-472a-91e8-745d895b4dff>
- Electronic Frontier Foundation (2021). UN Human Rights Committee Criticizes Germany's NetzDG for Letting Social Media Platforms Police Online Speech. Accessed via: <https://www.eff.org/deeplinks/2021/11/un-human-rights-committee-criticizes-germanys-netzdg-letting-social-media>
- Ermoshina, Ksenia (2022). Safe space by design? Federated architectures and alternative socio-technical models for content moderation, Alternative content regulation on social media. Accessed via: <https://carism.u-paris2.fr/fr/pre-conference-ica>
- Ertzscheid, Olivier, (2017). L'appétit des géants. Pouvoir des algorithmes, ambitions des plateformes, C&F éd.
- European Federation of Journalists (1971). Munich Declaration of the Duties and Rights of Journalists. Accessed via: <https://training.rsf.org/appendix-iii-declaration-of-rights-and-obligations-of-journalists/>
- European Liberal Forum (2022). The new Digital Service Act and the protection of freedom and rights in the digital public square. Accessed via: <https://liberalforum.eu/2022/01/digitising-europe-the-new-digital-service-act-and-the-protection-of-freedom-and-rights-in-the-digital-public-square/>
- European Parliament (2022). Digital Services Act. Accessed via: [https://www.europarl.europa.eu/doceo/document/TA-9-2022-0269\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-9-2022-0269_EN.html)
- European Parliamentary Research Service (2021). Liability of online platforms. Accessed via: [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS\\_STU\(2021\)656318\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU(2021)656318_EN.pdf)
- European Union (2019), Directive (EU) 2019/790. Accessed via: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L0790>
- European Union (2000), Directive 2000/31/EC. Accessed via: <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32000L0031>
- Fagan, Frank, (2018). Systemic social media regulation? Duke Law & Technology Review, Vol. 16, pp. 393. Accessed via: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3160781](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160781)
- Forum on Information & Democracy: Working group on infodemics (2020). How to end infodemics. Accessed via: [https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID\\_Report-on-infodemics\\_101120.pdf](https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf)
- Forum on Information & Democracy: Working group on the sustainability of journalism (2021). A New Deal for journalism. Accessed via: [https://informationdemocracy.org/wp-content/uploads/2021/06/ForumID\\_New-Deal-for-Journalism\\_16Jun21.pdf](https://informationdemocracy.org/wp-content/uploads/2021/06/ForumID_New-Deal-for-Journalism_16Jun21.pdf)
- Fukuyama, Francis, Barak Richman & Ashish Goel (2021). How to Save Democracy From Technology. Accessed via: <https://www.foreignaffairs.com/articles/usa/2020-11-24/fukuyama-how-save-democracy-technology>
- Gillespie, Tarleton, (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, Yale University Press. Accessed via: <https://yalebooks.yale.edu/book/9780300173130/custodians-internet>
- GPPI (2017). Internet Companies Cannot Be Judges of Free Speech. Accessed via: <https://www.gppi.net/2017/04/20/internet-companies-cant-be-judges-of-free-speech>
- Grimmelmann, James, (2019). Internet Law: cases & problems, Semaphore Press.
- G'sell, Florence, (2021). Le conseil de surveillance de Facebook et l'affaire Trump les limites du concept de 'tribunal d'entreprise'. La semaine du droit, éd. générale. n22°.
- Guéniat, Marc, (2020). Les géants du numérique deviennent les arbitres de la liberté d'expression. Accessed via: <https://www.heidi.news/innovation-solutions/les-reseaux-sociaux-mettent-les-internautes-dans-des-bulles-d-enfermement-culturelles>
- Hale, Thomas, David Held (2011). Handbook of Transnational Governance Institutions & Innovations, Polity Press.
- Haupt, Claudia E, (2021). Regulating Speech Online: Free Speech Values in Constitutional Frames, Washington University Law Review. Accessed via: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3794884#](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794884#)
- Hausalter, Louis, (2021). Liberté d'expression et pluralisme: face à Twitter et Facebook, l'impuissance des politiques. Accessed via: <https://www.marianne.net/politique/liberte-d-expression-et-pluralisme-face-a-twitter-et-facebook-l-impuissance-des-politiques>
- Heldt, Amélie & Stephan Dreyer (2021). Competent Third Parties and Content Moderation on Platforms: Potentials of Independent Decision-Making Bodies From A Governance Structure Perspective, Journal of Information Policy, Vol. 11, pp. 266-300. Accessed via: [https://leibniz-hbi.de/uploads/media/default/cms/media/6134zt1\\_Heldt-Amelie\\_jinfopoli.11.2021.0266.pdf](https://leibniz-hbi.de/uploads/media/default/cms/media/6134zt1_Heldt-Amelie_jinfopoli.11.2021.0266.pdf)

Heinrich-Böll-Stiftung (2019). Tackling hate speech requires the protection of freedom of expression. Accessed via: <https://ng.boell.org/en/2019/12/17/tackling-hate-speech-requires-protection-freedom-expression>

Hennebel, Ludovic, H  l  ne Tigroudja (2018). Trait   de droit international des droits de l'homme, Pedone.

Hirgen, Bezhani, (2021). Online Platforms and Freedom of Expression: The EU's Digital Services Act. Accessed via: <https://www.duo.uio.no/handle/10852/89284>

Hoog, Emmanuel, (2019). Confiance et libert   - Vers la cr  ation d'une instance d'autor  gulation et de m  diation de l'information. Accessed via: <https://www.culture.gouv.fr/Espace-documentation/Rapports/Confiance-et-liberte-Vers-la-creation-d-une-instance-d-autoregulation-et-de-mediation-de-l-information>

Horwitz, Jeff (2021). Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt, The Wall Street Journal. Accessed via: <https://www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353>

Indian government (2020). Central government Act, Section 79 in the Information Technology Act.

Institut Montaigne (2020). Disinformation: The German Approach and What to Learn From It. Accessed via: <https://www.institutmontaigne.org/en/blog/disinformation-german-approach-and-what-learn-it>

Institut Montaigne (2021). Internet: le p  ril jeune ? Accessed via: [https://www.institutmontaigne.org/ressources/pdfs/publications/rapport-internet-le-peril-jeune\\_0.pdf](https://www.institutmontaigne.org/ressources/pdfs/publications/rapport-internet-le-peril-jeune_0.pdf)

Institute for Governance and Policy Studies (2021). Regulating Harmful Communication: Current legal frameworks. After Christchurch: Hate, harm and the limits of censorship working paper series. Accessed via: [https://www.wgtn.ac.nz/\\_data/assets/pdf\\_file/0008/1926593/WP-21-05-regulating-harmful-communication-legal-frameworks.pdf](https://www.wgtn.ac.nz/_data/assets/pdf_file/0008/1926593/WP-21-05-regulating-harmful-communication-legal-frameworks.pdf)

Inter-American Dialogue (2021). Content Moderation and Self-Regulation Mechanisms: The Facebook Oversight Board and Its Implications for Latin America. Accessed via: <https://www.foreignaffairs.com/reviews/capsule-review/2021-12-14/content-moderation-and-self-regulation-mechanisms-facebook>

Isaac, Henri & Louis-Victor de Franssu (2022). Les utilisateurs doivent   tre int  gr  s    la r  gulation num  rique, Acteurs publics. Accessed via: <https://acteurspublics.fr/articles/les-utilisateurs-doivent-etre-integres-a-la-regulation-numerique>

Isaac, Stanley-Becker, Elizabeth Dvoskin (2020). Trump allies, largely unconstrained by Facebook's rules against repeated falsehoods, cement pre-election dominance. Accessed via: <https://www.washingtonpost.com/technology/2020/11/01/facebook-election-misinformation/>

Jourov  , V  ra (2021). Apr  s Trump, il est urgent d'encadrer les g  ants du num  rique, Lib  ration. Accessed via: [https://www.liberation.fr/debats/2021/01/22/apres-trump-il-est-urgent-d-encadrer-les-geants-du-numerique\\_1818153](https://www.liberation.fr/debats/2021/01/22/apres-trump-il-est-urgent-d-encadrer-les-geants-du-numerique_1818153)

Jurcys, Paulius, Christopher Donewald, Mark Fenwick, Markus Lampinen, Vytautas Nekro  sius, & Andrius Smaliukas (2020). Ownership of User-Held Data: Why Property Law is the Right Approach, SSRN Electronic Journal. Accessed via: <https://jolt.law.harvard.edu/assets/digestimages/Paulius-Jurcys-Feb-19-article-PJ.pdf>

Kaspar, Lea, Richard Wingfield (2021). Digital Rights at a Crossroads - recommendations for advancing human rights and social justice in the post-2020 era. Accessed via: <https://www.gp-digital.org/wp-content/uploads/2021/03/Digital-Rights-at-a-Crossroads.pdf>

Kaye, David, (2019). Speech Police, Columbia Global Reports.

Khan, Irene, (2021). Disinformation and freedom of opinion and expression Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. Accessed via: [https://eos.cartercenter.org/uploads/document\\_file/path/985/Report\\_of\\_the\\_Special\\_Rapporteur\\_on\\_Disinformation\\_and\\_Freedom\\_of\\_Opinion\\_and\\_Expression\\_E.pdf](https://eos.cartercenter.org/uploads/document_file/path/985/Report_of_the_Special_Rapporteur_on_Disinformation_and_Freedom_of_Opinion_and_Expression_E.pdf)

Khana, Tanveer, Antonis Michalassab & Adnan Akhuzadac (2021). Fake news outbreak 2021: Can we stop the viral spread? Accessed via: <https://www.sciencedirect.com/science/article/pii/S1084804521001326>

Lefouilli, Yassine & Leonardo Madio (2021). The Economics of Platform Liability. Accessed via: [https://www.tse-fr.eu/sites/default/files/TSE/documents/sem2021/eco\\_platforms/madio.pdf](https://www.tse-fr.eu/sites/default/files/TSE/documents/sem2021/eco_platforms/madio.pdf)

Legros, Claire, (2021). La libert   d'expression    l'heure du num  rique ou la difficile qu  te de l'  quilibre sur les r  seaux sociaux. Accessed via: [https://www.lemonde.fr/idees/article/2021/04/02/reseaux-sociaux-et-liberte-d-expression-inventer-des-dispositifs-pour-protger-nos-democraties\\_6075320\\_3232.html](https://www.lemonde.fr/idees/article/2021/04/02/reseaux-sociaux-et-liberte-d-expression-inventer-des-dispositifs-pour-protger-nos-democraties_6075320_3232.html)

Lingib  , Patrick, (2018). Quelles responsabilit  s sur les r  seaux sociaux ? Accessed via: <https://www.village-justice.com/articles/quelles-responsabilites-sur-les-reseaux-sociaux,29033.html>

Loutrel, Beno  t , Pierre Dubreuil & Sacha Desmaris (2019). Ce que peut l'Etat face aux plateformes. Accessed via: <https://www.numerique.gouv.fr/uploads/rapport-mission-regulation-reseaux-sociaux.pdf>

Mercatus center (2020). Do Our Leaders Believe in Free Speech and Online Freedom Anymore? Accessed via: <https://www.mercatus.org/bridge/commentary/do-our-leaders-believe-free-speech-and-online-freedom-anymore>

Musiani, Francesca, (2013). Governance by algorithm, Internet Policy Review, Vol.2. Accessed via: <https://doi.org/10.14763/2013.3.188>

Natalegawa, Andreyka & Kyra Jasper (2021). Controlling the Information Space: Big Tech and Free Speech in Southeast Asia, CSIS. Accessed via: <https://www.csis.org/blogs/new-perspectives-asia/controlling-information-space-big-tech-and-free-speech-southeast-asia>

Ndior, Val  re, (2022). Les r  seaux sociaux et l'administration Trump, H  rodote, Vol. 184-185, pp. 201    215. Accessed via: <https://www.cairn.info/revue-herodote-2022-1-page-201.htm>

Nicholas, Gabriel, (2022). Shedding Light on Shadowbanning, Center for Democracy & Technology. Accessed via: <https://cdt.org/wp-content/uploads/2022/04/remediated-final-shadowbanning-final-050322-upd-ref.pdf>

Observacom (2022). Twitter suspendi   la cuenta un portal ecuatoriano de periodismo de investigaci  n, luego de una publicaci  n vinculada al narcotr  fico. Accessed via: <https://www.observacom.org/twitter-suspendio-la-cuenta-un-portal-ecuatoriano-de-periodismo-de-investigacion-luego-de-una-publicacion-vinculada-al-narcotrafico/>

Observatoire européen de l'audiovisuel (2021). Décrypter la législation sur les services numériques. Accessed via: <https://rm.coe.int/iris-special-2021-01fr-decrypter-la-legislation-sur-les-services-numer/1680a4d2ba>

OCDE (2021). Transparency reporting on terrorist and violent extremist content online: An update on the global top 50 content sharing services. Éditions OCDE. Accessed via: <https://doi.org/10.1787/8af4ab29-en>

OFCOM (2021). Intermédiaires et plateformes de communication - Effets sur la communication publique et approches de gouvernance. Accessed via: <https://www.bakom.admin.ch/bakom/fr/page-daccueil/suisse-numerique-et-internet/communication-numerique/plate-formes-de-communication.html>

Office of the United Nations High Commissioner for Human Rights (1966). International Covenant on Civil and Political Rights. Accessed via: <https://www.ohchr.org/sites/default/files/Documents/ProfessionalInterest/ccpr.pdf>

Oversight Board (2022). Interview Quasi-Judicial Oversight Mechanisms for Social Platforms. Accessed via: <https://www.nomos-elibrary.de/10.5771/2699-1284-2021-3-254/quasi-judicial-oversight-mechanisms-for-social-platforms-a-conversation-with-catalina-botero-mari-no-co-chair-of-the-oversight-board-volume-2-2021-issue-3?page=1>

Pew Research Center (2017). The Future of Free Speech, Trolls, Anonymity and Fake News Online . Accessed via: <https://www.pewresearch.org/internet/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/>

Pisillo-Mazzeschi, Riccardo (2008). Responsabilité de l'État pour violation des obligations positives relatives aux droits de l'homme, Collected Courses of the Hague Academy of International Law, Vol. 333. Accessed via: [http://dx.doi.org/10.1163/1875-8096\\_pplrdc\\_A9789004172845\\_02](http://dx.doi.org/10.1163/1875-8096_pplrdc_A9789004172845_02)

Reisach, Ulrike (2021). The responsibility of social media in times of societal and political manipulation. European Journal of Operational Research. Accessed via: <https://www.sciencedirect.com/science/article/pii/S0377221720308249>

Reporters Without Borders (2016). RSF deplores suspension of French journalist's Facebook account. Accessed via: <https://rsf.org/en/rsf-deplores-suspension-french-journalist-s-facebook-account>

Reporters Without Borders (2021). India: RSF asks Twitter to restore The Kashmir Walla's account. Accessed via: <https://rsf.org/en/india-rsf-asks-twitter-restore-kashmir-walla-s-account>

Rest of world NY media (2021). How social media companies help African governments abuse "disinformation laws" to target critics. Accessed via: <https://restofworld.org/2021/social-media-africa-democracy/>

Revue internationale de droit économique (2019). Plateformes digitales et États: la corégulation par les données. Le cas des requêtes gouvernementales. Revue internationale de droit économique, Vol.33, pp. 363-381. Accessed via: <https://www.cairn.info/revue-internationale-de-droit-economique-2019-3-page-363.htm>

Rikke, F.J. & Lumi Zuleta (2020). Private Governance of Freedom of Expression on Social Media Platforms: EU content regulation through the lens of human rights standards. Nordicom Review, Vol. 41, pp. 51-67. Accessed via: <https://www.proquest.com/openview/a07f1352f6804ab21b66ee2bd5cd2e9e/1?pq-origsite=gscholar&cbl=2040374>

Pen America (2021). No Excuse No Abuse: What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users. Accessed via: <https://pen.org/report/no-excuse-for-abuse/>

Sander, Barrie (2020). Freedom of Expression in the Age of Online Platforms: the Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation. Fordham International Law Journal, Vol. 43, pp. 984-985

Stella, Elie, (2019). L'adaptation du droit pénal aux réseaux sociaux en ligne. Accessed via: <https://hal.univ-lorraine.fr/tel-02985468/document>

Strowel, Alain & Laura Somaini (2021). Towards a robust framework for algorithmic transparency to tackle the dissemination of illegal and harmful content on online platforms: Proposals for the Digital Services Act and beyond. Accessed via: [https://cdn.uclouvain.be/groups/cms-editors-crides/droit-intellectuel/CRIDES\\_WP\\_2\\_2021\\_Alain%20Strowel%20and%20Laura%20Somaini.pdf](https://cdn.uclouvain.be/groups/cms-editors-crides/droit-intellectuel/CRIDES_WP_2_2021_Alain%20Strowel%20and%20Laura%20Somaini.pdf)

Sudre, Frédéric, Hélène Surrel, Laure Milano & Béatrice Pastre-Belda (2021). Droit européen et international des droits de l'homme, PUF. .

Swedish Defence Research Agency (2019). Supervised Classification of Twitter Accounts Based on Textual Content of Tweets. Accessed via: [https://www.foi.se/download/18.46c3b221170ccb3336f1c6/1585826754727/Supervised-classification-of-twitter\\_FOI-S--6118--SE.pdf](https://www.foi.se/download/18.46c3b221170ccb3336f1c6/1585826754727/Supervised-classification-of-twitter_FOI-S--6118--SE.pdf)

Moore, Martin, and Tambini, Damian, (eds.) (2021). Regulating Big Tech: Policy Responses to Digital Dominance, Oxford University Press.

Timberg Craig, Elizabeth Dvoskin & Reed Albergotti (2021). Inside Facebook, Jan. 6 violence fueled anger, regret over missed warning signs. Accessed via: <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>

The Heartland Institute (2021). Six Principles for State Legislators Seeking to Protect Free Speech on Social Media Platforms. Accessed via: <https://www.heartland.org/publications-resources/publications/six-principles-for-state-legislators-seeking-to-protect-free-speech-on-social-media-platforms>

The Future of Free Speech (2021). A framework of first reference: Decoding a human rights approach to content moderation in the era of "platformization". Accessed via: [https://futurefreespeech.com/wp-content/uploads/2021/11/Report\\_A-framework-of-first-reference.pdf](https://futurefreespeech.com/wp-content/uploads/2021/11/Report_A-framework-of-first-reference.pdf)

The Future of Free Speech (2021). Rushing to Judgment: Are Short Mandatory Takedown Limits for Online Hate Speech Compatible with The Freedom of Expression? Accessed via: [https://futurefreespeech.com/wp-content/uploads/2021/01/FFS\\_Rushing-to-Judgment-3.pdf](https://futurefreespeech.com/wp-content/uploads/2021/01/FFS_Rushing-to-Judgment-3.pdf)

The Paradigm Initiative (2020). Digital rights and data privacy in Nigeria. Accessed via: [https://ng.boell.org/sites/default/files/2020-08/Digital%20Rights%20and%20Privacy%20in%20Nigeria\\_0.pdf](https://ng.boell.org/sites/default/files/2020-08/Digital%20Rights%20and%20Privacy%20in%20Nigeria_0.pdf)

The Santa Clara principles (2018). The Santa Clara Principles: On Transparency and Accountability in Content Moderation. Accessed via: <https://santaclaraprinciples.org/>

The World Wide Web Consortium (2022). Privacy Principles. Accessed via: <https://www.w3.org/TR/2022/DNOTE-privacy-principles-20220512/>

Tony Blair Institute (2021). Social Media Futures: What Is Brigading? Accessed via: <https://institute.global/policy/social-media-futures-what-brigading>

Tsamados, Andreas, Nikita Aggarwal, Josh Cows, Jessica Morley, Huw Roberts, Mariarosaria Taddeo & Luciano Floridi (2022). The ethics of algorithms: key problems and solutions. *AI & Soc*, Vol. 37, pp. 215-230. Accessed via: <https://doi.org/10.1007/s00146-021-01154-8>

Twitter (2019). World Leaders on Twitter: Principles & Approach. Accessed via: [https://blog.twitter.com/en\\_us/topics/company/2019/worldleaders2019](https://blog.twitter.com/en_us/topics/company/2019/worldleaders2019)

UN Human Rights Council (2018). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. Accessed via: <https://digitallibrary.un.org/record/1631686/usage>

UNESCO (2018). Journalism, 'Fake News' and Disinformation: Handbook for Journalism Education and Training. Accessed via: [https://unesdoc.unesco.org/ark:/48223/pf0000265552\\_eng](https://unesdoc.unesco.org/ark:/48223/pf0000265552_eng)

United Nations (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. Accessed via: [https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf)

United Nations (1969). Vienna Convention on the Law of Treaties. Accessed via: [https://legal.un.org/ilc/texts/instruments/english/conventions/1\\_1\\_1969.pdf](https://legal.un.org/ilc/texts/instruments/english/conventions/1_1_1969.pdf)

United States Congress (2020). Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2019, S.3398, 116th. Accessed via: <https://www.congress.gov/bill/116th-congress/senate-bill/3398/text>

Vengattil, Munsif & Elizabeth Culliford (2022). Facebook allows war posts urging violence against Russian invaders, Reuters. Accessed via: <https://www.reuters.com/world/europe/exclusive-facebook-instagram-temporarily-allow-calls-violence-against-russians-2022-03-10/>

Vermeule, Adrian & Cass Sunstein (2009). Conspiracy Theories: Causes and Cures. *The Journal of Political Philosophy*, Vol.17, pp. 202-227. Accessed via: <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9760.2008.00325.x>

Wong, Cynthia & James X. Dempsey (2011). The media and liability for content on internet. Accessed via: <https://www.opensocietyfoundations.org/uploads/ab948019-f717-436d-bdb8-b45893e7af08/mapping-digital-media-liability-content-internet-20110926.pdf>

Supported by



# Forum on Information & Democracy

Contact: [contact@informationdemocracy.org](mailto:contact@informationdemocracy.org)