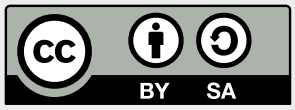
unesco

Elections in Digital Times_

A Guide for Electoral Practitioners

Published in 2022 by the United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France



Contributing Authors:
Prof. Dr. Dr. Robert Krimmer, Managing Director, E-Voting.CC GmbH, Competence Center for Electronic Voting and Participation, Professor of e-Governance and Digital Public Services, Center for IT Impact Studies, Johan Skytte Institute of Political Studies, University of Tartu
Dr. Armin Rabitsch, LL.M, Chairperson of Election-Watch.EU, International Monitoring and Missions Adviser of the University of Innsbruck, Peace and Conflict Studies
Rast'o Kužel, International Media and Election expert and Executive Director of MEMO 98.
Dr. Marta Achler, International Human Rights Lawyer, former Deputy Head of the Democratization Department of the OSCE /ODIHR
Nathan Licht, Consultant, E-Voting.CC GmbH, Competence Center for Electronic Voting and Participation

Editorial coordination:
Mehdi Benchelah, Senior Project Officer, Freedom of Expression and Safety of Journalists Section, UNESCO
Albertina Piterbarg, Consultant of Freedom, Expression and Safety of Journalists Section, UNESCO
Andrea Cairola, Programme Specialist, Freedom of Expression and Safety of Journalists Section, UNESCO

Graphic: Marcelo Falciani
Cover design: Marcelo Falciani

This publication was supported by the Multi-Donor Programme (MDP) on Freedom of Expression and Safety of Journalists.

With the support of the
UNESCO Multi-Donor Programme on Freedom of Expression
and Safety of Journalists (MDP)

Printed in France.

SHORT SUMMARY

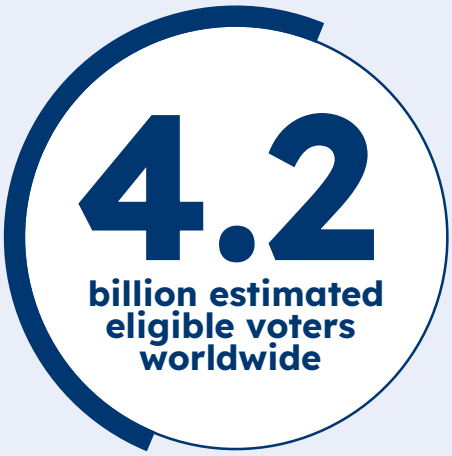
Strengthening democracy and electoral processes in the era of social media and Artificial Intelligence

Democracy requires free, periodic, transparent, and inclusive elections. Freedom of expression, freedom of the press, and the right to political participation are also critical to societies ruled by the respect of human rights. In today's rapidly evolving digital environment, opportunities for communication between citizens, politicians and political parties are unprecedented — with information related to elections flowing faster and easier than ever, coupled with expanded opportunities for its verification and correction by a growing number of stakeholders. However, with billions of human beings connected, and disinformation and misinformation circulating unhinged around the networks, democratic processes and access to reliable information are at risk.

With an estimated 56.8% of the world's population active on social media and an estimate of 4 billion eligible voters, the ubiquity of social networks and the impact of Artificial Intelligence can intentionally or unintentionally undermine electoral processes, thereby delegitimizing democracies worldwide.

In this context, all actors involved in electoral processes have an essential role to play. Electoral management bodies, electoral practitioners, the media, voters, political parties, and civil society organizations must understand the scope and impact of social media and Artificial Intelligence in the electoral cycle. They also need to have access to the tools to identify who instigates and spreads disinformation and misinformation, and the tools and strategies to combat it.

This handbook aims to be a toolbox that helps better understand the current scenario and share experiences of good practices in different electoral settings and equip electoral practitioners and other key actors from all over the world to ensure the credibility of the democratic system in times of profound transformations.



Elections in Digital Times

A Guide for Electoral Practitioners

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF BOXES

EDITORIAL TEAM

Authors

- Prof. Dr. Dr. Robert Krimmer, Managing Director, E-Voting.CC GmbH, Competence Center for Electronic Voting and Participation, Professor of e-Governance and Digital Public Services, Center for IT Impact Studies, Johan Skytte Institute of Political Studies, University of Tartu
- Dr. Armin Rabitsch, LL.M, Chairperson of Election-Watch.EU, International Monitoring and Missions Adviser of the University of Innsbruck, Peace and Conflict Studies
- Rast'o Kužel, International Media and Election expert and Executive Director of MEMO 98
- Dr. Marta Achler, International Human Rights Lawyer, former Deputy Head of the Democratization Department of the OSCE /ODIHR
- Nathan Licht, Consultant, E-Voting.CC GmbH, Competence Center for Electronic Voting and Participation

Editorial Coordination

- Mehdi Benchelah, Senior Project Officer of Freedom of Expression and Safety of Journalists Section, UNESCO
- Andrea Cairola, Programme Specialist, Freedom of Expression and Safety of Journalists Section, UNESCO
- Albertina Piterbarg, Consultant, Freedom of Expression and Safety of Journalists Section, UNESCO

Contributions

- Guilherme Canela, Chief, Freedom of Expression and Safety of Journalists Section, UNESCO
- Guy Berger, Director, Communication and Information Sector, Strategic Planning Office, UNESCO
- Macarena Rivera Lam, Consultant, Freedom of Expression and Safety of Journalists Section, UNESCO
- Klara Vandeborne, Trainee, Freedom of Expression and Safety of Journalists Section, UNESCO
- Rosario Soraide, Consultant, Freedom of Expression and Safety of Journalists Section, UNESCO
- Hanna Fiskesjö, Associate Programme Specialist, Communication and Information Sector, UNESCO
- Julie Ballington, Policy Advisor on Political Participation in the Leadership and Governance Section, UN Women
- Lana Ačkar, Policy Specialist, Women's Political Participation Leadership and Governance Section, UN Women

Advisory Board

International Organisations

- Olufunto Akinduro, Senior Programme Officer Elections, Africa Regional Programme, International IDEA
- Julia Brothers, Programme Director, National Democratic Institute (NDI)
- Patrick Costello, Head of Global 3 Division, European External Action Service (EEAS)
- Gerardo Icaza, Director of Electoral Cooperation and Observation, Organisation of American States (OAS)
- Idriss Kamara, Political and Electoral Affairs Officer, African Union (AU)
- Beata Martin-Rozumilowicz, Director for Europe and Eurasia, IFES
- Raymond Serrato, Open Source Investigator, UNHR
- Alex Shlyk, Head of Election Department, Office for Democratic Institutions and Human Rights (ODIHR), Organisation for Security and Co-operation in Europe (OSCE)
- Peter Wolf, Senior Expert, International IDEA

Non-Government Organisations and Industry

- Kala Fleming, CEO, Diaspora AI
- Rafael Goldzweig, Research Coordinator, Democracy Reporting International (DRI)
- Nouri Lajmi, President, High Authority for Audiovisual Communication (HAICA)
- Chandanie Watawala, Executive Director, Asian Network for Free Elections (ANFREL), Myanmar
- Giovanna Maiola, Research Associate, Osservatorio Di Pavia
- Julia Pomares, Executive Director, Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento (CIPPEC)
- Khadeja Ramali, Social Media Expert, Project Silphium

LIST OF ACRONYMS

ASEAN	The Association of Southeast Asian Nations
DOS	Denial-of-service attack
DDOS	Distributed denial-of-service attack
ECHR	European Convention on Human Rights
ECtHR	European Court on Human Rights
EMB	Electoral Management Bodies
ENISA	European Union Agency for Network and Information Security
EU	European Union
ICCPR	International Covenant on Civil and Political Rights
ISP	Internet Service Providers
MIL	Media and information literacy
SDG	Sustainable Development Goals
COMEST	World Commission on the Ethics of Scientific Knowledge and Technology
OAS	Organization of American States
OHCHR	Office of the High Commissioner on Human Rights
OSCE	Organization for Security and Co-operation in Europe
UDHR	Universal Declaration of Human Rights
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNDP	United Nations Development Programme
VAW	Violence against women
VAWP	Violence against women in politics
VAWIE	Violence against women in elections
WPPRI	Women's Political Participation Risk Index

METHODOLOGICAL NOTE

This handbook offers a comprehensive overview of international and regional standards and commitments related to the rights to freedom of expression, access to information, political participation, and privacy in the area of the internet, social media and Artificial Intelligence in elections.

It also maps a series of good practices implemented by diverse stakeholders worldwide during electoral processes. It is organized in eight sections: six sections dedicated to the analysis of the challenges and the general situation of elections in digital times, a section dedicated to conclusions and a section with suggestions for possible action.

Each of the first six sections has a guide of suggested questions at the end in case the publication is used for trainings and workshops.

FOREWORD

The swift development of the internet, social media and Artificial Intelligence (AI) has profoundly impacted elections and democratic processes worldwide – with both benefits and drawbacks.

The digital era provides new, more direct and potentially transparent means for Election Management Bodies (EMBs) to monitor polls and interact with the electorate. Politicians and political parties can now communicate directly with their supporters and canvass voters on a large scale.

At the same time, these developments have also posed new risks to the credibility and integrity of democracies. UNESCO's 2019 report on Elections and media in digital times identified three global threats of significance for elections in today's increasingly interconnected societies: (1) the spread of disinformation and misinformation, as well as hate-speech; (2) the increase in intimidations and violence against journalists and media actors; and (3) disruptions in electoral campaigning and communications.

The challenge is to optimize the pros and minimize the cons linked to digital technologies, the companies providing these services and the users of these tools. The goal must be to strengthen the exercise of political rights and the transparency of electoral processes, rather than hamper or endanger them. This objective is relevant to electoral practitioners, EMBs, legislators, international institutions, civil society organizations, the media, politicians, security forces and the general public. How to achieve it? Existing international law and good practices, elaborated in this publication, can function as a valuable guide and benchmark.

The onset of the COVID-19 pandemic, which started in 2020, brought new challenges to managing elections worldwide. It also renewed discussions on the potential role of digital tools in facilitating voting. Notably, the pandemic also resulted in new insight regarding responses by social media platforms in moderating content that could imminently affect public health, and which has significance to elections. This resonates also with calls for faster action for social media platforms in dealing with political disinformation and hate speech.

The COVID-19 crisis also highlighted the importance of professional media, access to verified information, and increased support for fact-checking initiatives. Again, this echoes the important role of journalism in the context of elections, in particular in digital times.

UNESCO conceived this handbook to provide practical tools for a range of key electoral stakeholders in response to these pressing needs. It seeks to contribute to the achievement of Sustainable Development Goal (SDG) 16, which focuses on peace, justice, and strong institutions, and especially to SDG target 16.7 ("Ensure responsive, inclusive, participatory and representative decision-making at all levels") and SDG target 16.10 ("Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements").

The pages offer a comprehensive overview of international and regional standards and commitments related to the rights to freedom of expression, access to information, political participation, and privacy, which are vital when considering the impact of the internet, social media and AI on elections. It also maps a series of good practices implemented by diverse stakeholders worldwide.

Finally, this handbook outlines suggestions for possible action by diverse electoral practitioners who are at the frontline – serving as a practical toolkit for them.

The present publication is the result of a long-standing work of UNESCO in media and elections, which is in turn embedded within the UN's broader electoral assistance efforts. It was developed with the help of a panel of 16 experts from across regions, including experts involved in the fields of elections, technology, media and freedom of expression, among others. We thank them for their insights.

Dr. Tawfik Jelassi
UNESCO Assistant Director-General for Communication and Information

1. INTRODUCTION: ARTIFICIAL INTELLIGENCE AND ELECTIONS

OBJECTIVES OF THIS SECTION

- To understand the concept of Artificial Intelligence (AI) and the main issues arising from the new digital environment for democracies.
- To familiarize with how an algorithm works and how it could affect freedom of expression.
- To comprehend how social networks influence political discussion.
- To identify the challenges of AI all along the Electoral Cycle.

The field of Artificial Intelligence (AI) has been making breathtaking advances for the past two decades. In particular, it is contributing to the automation of data analysis. Artificial Intelligence is no longer programmed line by line, but is now capable of learning, thereby continuously developing itself.[1] And, while it has the potential to improve human existence, at the same time it threatens to deepen social divides.

The mechanics behind AI are quite straightforward: search engines and recommendation platforms identify personalised suggestions for products and services based on personal preferences and meta-data that has been gathered from previous searches, purchases and mobility behaviour, as well as social interactions. While officially, the identity of the user should be protected, it can, in practice, be inferred quite easily. Today, and thanks to machine learning, algorithms can increasingly predict people's preferences. But the more AI knows about the users and consumers, the less likely their choices are to be open and not predetermined by different agendas and political interests.

A further problem arises when adequate transparency and democratic controls are lacking. Search algorithms and recommendation systems can be influenced. Companies can bid on certain combinations of words to gain more favourable results. Governments are probably able to influence the outcomes too. During elections, political actors might nudge undecided voters towards supporting them. Therefore, whoever controls this technology would have an important advantage to win elections. This problem is exacerbated by the fact that, in many countries, a single search engine or social media platform has a predominant market share. It could decisively influence the public and interfere with other countries remotely too.[2]

In order for this engineering to stay unnoticed, it takes a so-called resonance effect—suggestions that are sufficiently customized to each individual. In this way, local trends are gradually reinforced by repetition, leading all the way to the «filter bubble» or «echo chamber effect»: in the end, all a person might get is their own opinions reflected at them. This might cause social polarization, resulting in the formation of separate groups that no longer understand each other and find themselves increasingly at conflict with one another. In this way, personalized information could hamper social cohesion.[3]

As the influence and impact of AI spread, it will be critical to involve people and experts from the most diverse backgrounds possible in guiding this technology in ways that enhance human capabilities and lead to positive outcomes. And this is critical for democratic institutions, electoral mechanisms, and political life in general. The dynamics between social media, AI and elections is complex, problematic, and full of tensions.

---

1 World Economic Forum, Strategic Intelligence: https://intelligence.weforum.org/.

2 D. Helbing, B. S. Frey, G. Gigerenzer, E. Hafen, M. Hagner, Y. Hofstetter, J. van den Hoven, R. V. Zicari and A. Zwitter, February 25, 2017, ***Will Democracy Survive Big Data and Artificial Intelligence?***, Scientific American, https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/.

3 Ibid.

In the same way, tensions between freedom of expression, the role of journalism, social media, and electoral processes are also aggravated by Artificial Intelligence mechanisms. Platforms may choose to increase traffic to privilege certain algorithms over others. For example, algorithms that prioritize the most controversial exchanges usually generate more likes and engagements for the platform and, therefore, make the platform's advertising more relevant. It does not matter if the news and information they distribute are legitimate or not. The important thing is to have a lot of user traffic.

The use of algorithms is critical but at the same time inaccessible to most governments, as it is unregulated, or only partially so, and remains in the hands of private parties with whom agreements must be reached in order to protect democracy and electoral mechanisms.

Understanding this dynamic is fundamental for electoral bodies and practitioners to safeguard the integrity and credibility of electoral processes, as well as the role of the news media all along the electoral cycle, in the face of new issues related to AI and the digital environment. This includes (i) online disinformation; (ii) the digital dimension of the safety of journalists and other media actors, and (iii) disruptive practices in election campaigning and communications.[4]

Regarding the first category, while disinformation as an escalating trend impacts several critical aspects (for example, public health) it is of particular significance concerning whether societies have informed electorates.

The second category includes the continued and digitally intensified patterns of threats and violence against journalists and other actors who contribute to public debate. Killings of journalists and impunity for killings remain at shocking levels. There is also a growing urgency about escalating threats and violence against women journalists. Rhetorical assaults, including by political actors, and the increasing digital dimension to attacks on journalists, are worrying trends in general, and with special relevance for elections.[5]

The third category concerns the digitally enabled disruption of elections and the news media's role in political communications. Disruption can take many forms, such as the circumvention of campaign financing rules; the lack of transparency in political advertising; the fragmentation of public space through political micro-targeting; ethical shortcomings by politicians, media and Internet actors during election periods; and political actors being able to bypass scrutiny by traditional media outlets and associated regulations to reach voters directly through Internet platforms. The key tasks of the media in any democratic society – to inform the public about matters of interest to society; to act as public watchdogs exposing corruption and wrongdoing; and to provide a shared forum for public debate – take on added importance in the context of elections.

The information and ideas disseminated and debated during election periods influence public opinion- and decision-making processes, which find ultimate and formal expression in the ballot box. Disruptive practices in relation to elections underscore the need for public debate to be nourished by accurate and reliable information. While each of these three themes has its own distinctive dynamics and drivers, the interplay between them in relation to elections is particularly powerful, as shown in the graphic below:

---

[4] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

[5] Violence against women politicians and women in public life (activists included) is also concerning. In the UN General Assembly Resolution A/RES/73/148 (2018) the UN General Assembly calls upon States to prevent, address and prohibit violence, including sexual harassment, against women and girls in public and political life, including women in leadership positions, journalists and other media workers and human rights defenders, including through practical steps to prevent threats, harassment and violence, including by combatting impunity and ensuring that those responsible for violations and abuses, including sexual and gender-based violence and threats, including in digital contexts, are promptly brought to justice and held accountable through impartial investigations. See also: https://unesdoc.unesco.org/ark:/48223/pf0000371524.

New developments highlight the need to safeguard the integrity of electoral processes, as well as the role of media during election periods:

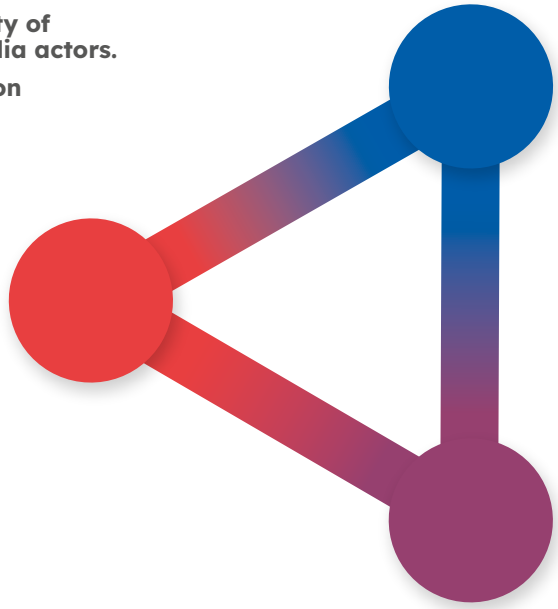
- **Disinformation and misinformation.**
- **Attacks on the safety of journalists and media actors.**
- **Disruption in election campaigning and communications.**

JOURNALISTS UNDER FIRE

Threats and violence against journalists have continued and expanded in recent years.

Killings of journalists and impunity for the killings remain at shocking levels.

Hostile rhetoric and online threats to media actors are a growing trend.



INFORMATION UNDER ATTACK

So called "fake news" has become a dominant term, but is also now experiencing push back.

Disinformation and **misinformation** have emerged as preferred ways to describe content that undermines accuracy and reliablity of information that underpins public opinion.

ELECTION INTEGRITY AT RISK

Disruption of democratic processes today includes:

- circumvention of campaign financing rules
- lack of transparency in political advertising and political micro-targeting
- crackdowns on legitimate political content
- shutdowns of internet access and applications.

FIGURE 1: THREE TRENDS CONVERGE[6]

The impact of new technologies throughout the electoral cycle implies the optimization of resources and the emergence of new challenges. In this sense, all actors involved in the electoral processes must understand the scope and impact of this landscape, be aware of the issues, and have the tools to implement solutions. Governments, electoral bodies, practitioners, political parties, candidates, journalists, religious and traditional leaders, civil society, and the general population must have access to education to understand this new era better and protect their democratic institutions.

The United Nations, in its holistic approach to elections, has already begun to provide through its support programs, technological assistance and training to strengthen the capacity of all key stakeholders in the new technological landscape. In 1991, the General Assembly established a framework for United Nations electoral assistance, which has continued to evolve and remains the basis for United Nations work in this field. The organization provides assistance only at the specific request of the Member State concerned or as mandated by the Security Council or General Assembly.

---

[6] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***.

BOX 1: THE UN APPROACH TO ELECTORAL ASSISTANCE

The United Nations plays a major role in providing international electoral assistance at the specific request of a Member State, concerned or based on a mandate from the Security Council or General Assembly. UN electoral support programs are tailored according to the specific needs of each requesting Member State and may include among other:

- **Technical Assistance:** Legal, operational and logistic assistance provided to develop or improve electoral laws, processes and institutions.
- **Support to creating a conducive environment:** The mandate of UN peace operations often includes provisions related to creating a conducive environment for elections.
- **Electoral Observation:** UN election observation entails the deployment of a mission to observe each phase of an electoral process and report back to the Secretary-General.
- **Panels of Political and/or Electoral Experts:** A panel can be an electoral expert monitoring team, composed of experts in such areas as electoral processes or mediation, or a high-level one composed of eminent persons of political, electoral or mediation profile.

UN electoral assistance is a system-wide endeavor, tapping the complementary expertise and capacities of many parts of the UN family. Entities providing electoral assistance usually are:

- **The Department of Political and Peacebuilding Affairs (DPPA) and the Inter-Agency Coordination Mechanism for Electoral Assistance (ICMEA)**

  The Under-Secretary-General for Political Affairs and head of DPPA serves as the United Nations Focal Point for electoral assistance and is supported in that function by DPPA's Electoral Assistance Division (EAD). Importantly, the EAD convenes and chairs the Inter-Agency Coordination Mechanism for Electoral Assistance (ICMEA), which is a platform that facilitates the information sharing, coordination and internal policy-development among UN entities that engage on electoral assistance.

- **The Department of Peace Operations (DPO)**

  In peacekeeping and many post-conflict environments, assistance is generally provided through electoral components of field missions under the aegis of the Department of Peace Operations.

- **The United Nations Development Programme (UNDP)**

  UNDP is the major implementing body of the Organization for support to developing electoral institutions, legal frameworks and processes and support to elections outside the peacekeeping or post-conflict context. It manages some 40 to 50 electoral projects per year.

- **The United Nations Educational, Scientific and Cultural Organization (UNESCO)**

  UNESCO is the United Nations specialized agency tasked with promoting and supporting freedom of expression, press freedom and freedom of information. Free, independent media, online as well as offline, are essential to the integrity of electoral processes.

- **The Office of the High Commissioner for Human Rights (OHCHR)**

  OHCHR provides training and advice on human rights monitoring in the context of elections, supports and organizes campaigns for violence-free elections, engages in

advocacy for human rights-compliant electoral laws and institutions, and monitors and reports on human rights violations during electoral processes.

- **The United Nations Entity for Gender Equality and the Empowerment of Women (UN-Women**

  UN-Women is mandated to provide, through its normative support functions and operational activities, guidance and technical support to all Member States, at their request, on gender equality, the empowerment and rights of women and gender mainstreaming. It promotes gender equality and women's participation in political processes.

- **The International Organization for Migration (IOM)**

  IOM which joined the United Nations system in 2016, is the leading intergovernmental organization in the field of migration and often implements out-of-country voting programmes for refugees, asylum seekers and migrants.

For more information on how the United Nations provides electoral support, see: https://dppa.un.org/en/elections.

1.1. ARTIFICIAL INTELLIGENCE AND SOCIAL MEDIA

Artificial Intelligence (AI) has been developing rapidly in recent decades yet lacks a universally accepted definition.[7] While there is no one single agreed definition, this Guide will focus on the combination of technologies that UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) has described as "machines capable of imitating certain functionalities of human intelligence, including such features as perception, learning, reasoning, problem-solving, language interaction, and even producing creative work".[8]

On 24 November 2021, the Recommendation on the Ethics of Artificial Intelligence was adopted by UNESCO's General Conference at its 41st session.[9] UNESCO embarked on a two-year process to elaborate this first global standard-setting instrument on the ethics of AI in the form of a Recommendation, following the decision of its General Conference at its 40th session in November 2019. UNESCO Recommendation approached AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control.

In current AI research, it is mostly the aspect of rationality that is considered decisive for a machine to classify as intelligent.[10] Rationality refers to the machine's ability to perceive its environment by collecting and interpreting data and, furthermore, the ability to apply reasoning to collected data

---

[7] J. Berryhill, K.K. Heang, R. Clogher and K. McBride, 2019, ***Hello, World: Artificial intelligence and its use in the public sector***, OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, https://doi.org/10.1787/726fd39d-en.

[8] As cited in X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019, ***Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective***, UNESCO, p. 24.

[9] ***Recommendation on the Ethics of Artificial Intelligence*** adopted by UNESCO's General Conference at its 41st session, 24 November 202. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[10] EC (Venice Commission), 2019, ***Joint Report of the Venice Commission and of the Directorate of Information Society and Action against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections***, CDL-AD (2019)016.

and hence decide on how to best react to it.[11] This consideration holds for a large part of what AI technology is used for in elections, such as the collection and processing of information that is transformed into electoral advertisement content.[12]

The European Commission for Democracy through Law (Venice Commission)[13] has defined social media as: "The web or mobile-based platforms that allow for two-way interactions through user-generated content (UGC) and communication. Social media are therefore not media that originate only from one source or are broadcast from a static website. Rather, they are media on specific platforms designed to allow users to create ("generate") content and to interact with the information and its source. While social media rely on the Internet as a medium, it is important to note that not all Internet sites or platforms meet the definition of social media. Some websites make no provision for interactivity with the audience, while others allow users only to post comments as a reaction to particular published content as discussions posts (or 'threads') which are moderated and controlled. While discussion threads can offer a degree of interaction with the source, these are not considered to be social media platforms".[14]

Social media has become an integral part of our societies, and thus it plays an essential role in democracies and electoral processes. Globally some 58.4 percent of the population (more than 3 billion users) use social media networks, with Facebook being the biggest one, having around 2.910 billion monthly active users.[15] Several services these companies offer are social and have a personal character, such as closed groups on Facebook and messaging applications.

AI systems raise new types of ethical issues that include, but are not limited to, their impact on decision-making, employment and labour, social interaction, health care, education, media, access to information, digital divide, personal data and consumer protection, environment, democracy, rule of law, security and policing, dual use, human rights and fundamental freedoms, including freedom of expression, privacy and non-discrimination. Furthermore, new ethical challenges are created by the potential of AI algorithms to reproduce and reinforce existing biases, and thus to exacerbate already existing forms of discrimination, prejudice and stereotyping.

Some of these issues are related to the capacity of AI systems to perform tasks which previously only living beings could do, and which were in some cases even limited to human beings only. These characteristics give AI systems a profound, new role in human practices and society, as well as in their relationship with the environment and ecosystems, creating a new context for children and young people to grow up in, develop an understanding of the world and themselves, critically understand media and information, and learn to make decisions. In the long term, AI systems could challenge humans' special sense of experience and agency, raising additional concerns about, inter alia, human self-understanding, social, cultural and environmental interaction, autonomy, agency, worth and dignity.

1.2. THE ELECTORAL CYCLE APPROACH

Elections are composed of a number of integrated building blocks, with different stakeholders interacting and influencing each other. Electoral components and stakeholders do not stand alone. They are interdependent, and therefore the breakdown of one aspect (for example the collapse of

---

[11] Ibid.

[12] S. Monteleone, 2019, ***Artificial intelligence, data protection and elections***; F. Zuiderveen Borgesius, J. Möller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodo and C. de Vreese, 2018, Online Political Microtargeting: Promises and Threats for Democracy, ***Utrecht Law Review***, 14(1).

[13] EC (Venice Commission), 2019, p. 3.

[14] S. Kaiser, 2014, ***Social Media A Practical Guide for Electoral Management Bodies***, International Institute for Democracy and Electoral Assistance (IDEA), p. 11. Available at: https://www.idea.int/sites/default/files/publications/social-media-guide-for-electoral-management-bodies.pdf.

[15] See: https://datareportal.com/reports/digital-2022-global-overview-report.

a particular system of voter registration) can negatively impact every other, including human and financial resources, the availability of supplies, costs, transport, training and security, and thus on the credibility of the election itself. In turn, if an electoral process suffers from low credibility, this is likely to damage the democracy and its institutions.

The cyclical approach is a key instrument to facilitate the understanding of the interdependence of different electoral activities, helping Election Management Bodies (EMBs) officials, electoral practitioners and donors to plan and allocate resources for specific activities in a timelier fashion than in the past.[16] In particular, it places an important emphasis on the post-electoral period as a significant moment of institutional growth as opposed to the vacuum between elections.

Moreover, the electoral cycle approach supports development agencies and partner countries to plan and implement electoral assistance within the democratic governance framework by thinking ahead, rather than reacting to each electoral event as it occurs.[17]

From the perspective of an EMB, the electoral cycle encompasses all steps and processes that fall within the extent of its functions, responsibilities, and powers that are necessary for an election or vote to take place and assists in the strategic and operational planning.

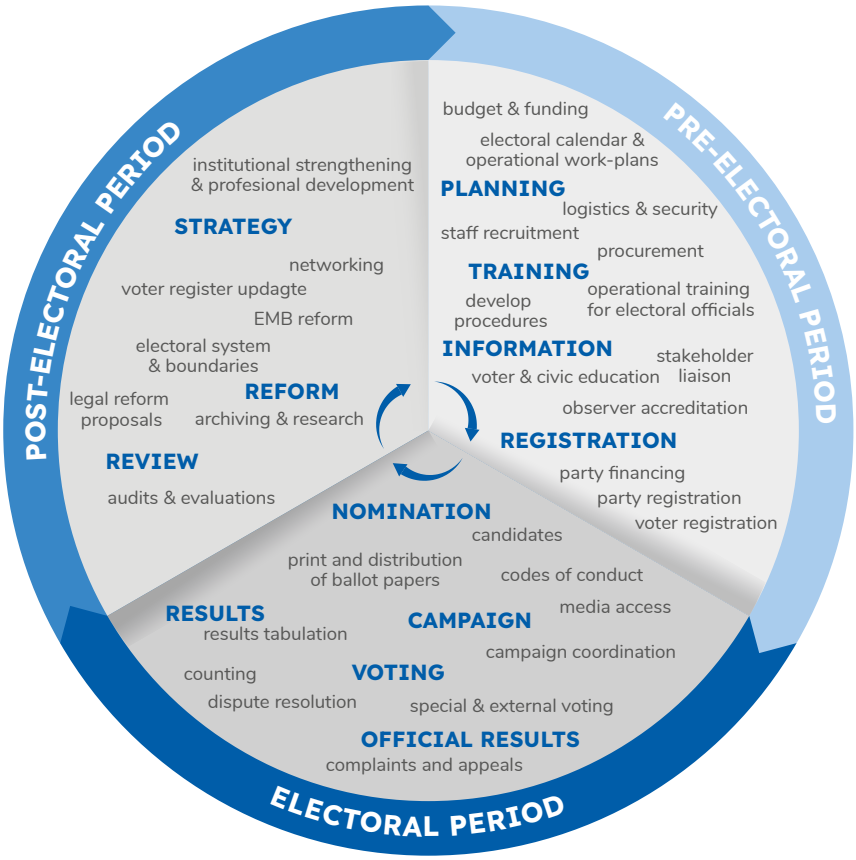


FIGURE 2: THE ELECTORAL CYCLE

[16] United Nations A/RES/60/162, General Assembly Distr.: General 28 February 2006 Sixtieth session Agenda item 71 (b) 05-49732, Resolution adopted by the General Assembly on 16 December 2005. Available at: https://digitallibrary.un.org/record/563281?ln=en#record-files-collapse-header.

[17] For more information see The ACE Electoral Knowledge Network, **"What is the Electoral Cycle?"**. Available at: https://aceproject.org/electoral-advice/electoral-assistance/electoral-cycle.

Developing a strategic plan is a basic step in focusing on EMB's efforts on achieving a set of agreed objectives based on its legally defined responsibilities and the different phases of the electoral cycle. All electoral practitioners and main actors involved in the process – including the media, for their planning purposes should follow this approach, taking into account the technological challenges imposed on each stage of the cycle by new technologies, social media and AI.

ACTIVITY I

The following activity has the objective of determining the reader's/participant's level of knowledge of the mechanics behind AI and social media, the impact on democracies and electoral processes worldwide.

Suggested guiding questions for a discussion:

I. Please define Artificial Intelligence. According to your own experience, can you relate AI to your everyday life? How?
II. Can you explain how an algorithm works and how it could affect freedom of expression?
III. Do you believe that social networks influence political discussion in your country? Why?
IV. Which are the main issues arising from the new digital environment?
V. How could AI affect the Electoral Cycle? Do you have any examples?

2. INTERNATIONAL HUMAN RIGHTS LAW FRAMEWORK

OBJECTIVES OF THIS SECTION

- Examine the new challenges to Human Rights in the Digital Era.
- Provide an overview of the international human rights law framework and the international standards and soft law on freedom of expression and the right to privacy.
- Understand the impact of AI on Human Rights in the context of electoral processes.
- Set out the relevance of women's rights and political participation, freedom of expression, safety of journalists, hate speech and the main concerns regarding democratic accountability.

The Universal Declaration of Human Rights recognizes that "everyone has the right to take part in the government of [their] country, directly or through freely chosen representatives" and that "the will of the people shall be the basis of the authority of government; this shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures".[18] Normative frameworks also emphasize equal rights for women, ethnic, religious and linguistic minorities, Indigenous Peoples, youth and persons with disabilities.

International human rights instruments protect a number of basic rights, the enjoyment of which are crucial for a meaningful electoral process. Furthermore, the right to participate in genuine and periodic elections implies other rights, including: the right to freedom of expression, the right to freedom of opinion, the right to freedom of association, the right to peaceful assembly and the right to privacy. The rights individuals enjoy offline also apply online. These rights and principles are enshrined in the Charter of the United Nations, the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and other relevant documents.

2.1. NEW CHALLENGES TO HUMAN RIGHTS IN THE DIGITAL ERA

The Internet, social media, and AI pose challenges to the electoral processes and the implementation of fundamental human rights and internationally accepted standards and norms. Mainly they affect those human rights related to freedom of expression, the right of peaceful assembly, the safety of candidates, the right to political participation, the situation and rights of women in politics, activists, journalists, and bloggers, the right to privacy, and hate speech/ inflammatory language.

In her keynote speech during an event titled "Human rights in the digital age - Can they make a difference?" in October 2019,[19] Michelle Bachelet, UN High Commissioner for Human Rights, highlighted that the "digital revolution is major global human rights issue whose unquestionable benefits do not cancel out its unmistakable risks".[20]

Harassment, trolling campaigns and intimidation have polluted parts of the Internet and pose very real off-line threats, with a disproportionate impact on women.[21]

Threats, intimidation, and cyber-bullying on the Internet lead to real world targeting, harassment, violence, and murder, even to alleged genocide and ethnic cleansing. Failure to take action might

---

[18] See: https://datareportal.com/reports/digital-2022-global-overview-report.

[19] Keynote speech by Michelle Bachelet, UN High Commissioner for Human Rights, 17 October 2019, ***Human Rights in the Digital Age***, Japan Society, New York. Available at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25158.

[20] Ibid.

[21] Also see: UNESCO, 2021, ***Practical guide for women journalists on how to respond to online harassment***, Paris. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000379908.

result in further shrinking of civic space, decreased participation, enhanced discrimination, and a continuing risk of lethal consequences – in particular for women, minorities and migrants.

But over-reaction by regulators, often under the pretence of fighting hate or extremism, to rein in speech and use of the online space is also a critical human rights issue. Dozens of countries are limiting what people can access online, curbing free speech and political activity, often under the pretence of fighting hate or extremism. Internet shutdowns seem to have become a common tool to stifle legitimate debate, dissent and protests. The NGO Access Now counted at least 155 shutdowns in 29 countries in 2020.[22]

Some States are using digital surveillance tools to track down and target rights defenders and other people perceived as critics. States and businesses are already using data-driven tools that can identify individuals as potential security threats, including at borders and in criminal justice systems. According to Bachelet, Artificial Intelligence systems assess and categorize people; draw conclusions about their physical and mental characteristics; and predict their future medical conditions, their suitability for jobs, even their likelihood of offending.[23] People's profiles, "scoring" and "ranking" can be used to assess their eligibility for health care, insurance and financial services but also to be surveilled regarding their ideas and political participation.

So, alongside the human rights abuses, there is a whole new category not necessarily deliberate, not the result of a desire to control, but by-products of a drive for efficiency and progress. Real world inequalities are reproduced within algorithms and flow back into the real world. Artificial Intelligence systems cannot capture the complexity of human experience and need. People's data is not just digitized but monetized and politicized.[24]

This situation challenges the principles of the Universal Declaration of Human Rights. Each person is equal, an individual with inalienable rights and inherent dignity. Each person has the right to live their life free from discrimination to political participation, privacy, health, liberty, a fair trial. To respect these rights in this rapidly evolving world, it is critical to ensure that the digital revolution is serving the people, and that AI systems comply with cornerstone principles such as transparency, fairness, accountability, oversight and redress.[25]

To tackle these multiple, complex risks that cross cultures, national boundaries, and legal jurisdictions, a universal human response in defence of universal human rights is needed. The international human rights framework takes us further than ethics alone in placing the necessary checks and balances on this power. It provides a concrete, legal foundation on which States and firms can build their responses in the digital age and clear guidance on acceptable behaviour.

There are numerous conventions, treaties, courts, commissions, and other institutions that can hold States and companies to account. Alongside the UN Guiding Principles on Business and Human Rights,[26] there are already several examples of guidance in specific sectors, such as the European Union's ICT Sector Guidance on implementing the Guiding Principles,[27] the UNESCO Recommendation on the Ethics of AI,[28] the Telecommunications Industry Dialogue and the GNI Principles and Guidelines that look to keep the power of data-driven companies and States in check.

---

[22] See: https://www.accessnow.org/; #KeepItOn: Fighting internet shutdowns around the world (accessnow.org).

[23] Michelle Bachelet, UN High Commissioner for Human Rights, 17 October 2019.

[24] Ibid.

[25] Ibid.

[26] OHCHR, 2011, ***UN Guiding Principles on Business and Human Rights***. Available at: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

[27] European Union, 2011, ***European Union's ICT Sector Guidance on implementing the Guiding Principles***. Available at: https://www.ihrb.org/pdf/eu-sector-guidance/EC-Guides/ICT/EC-Guide_ICT.pdf.

[28] See: https://unesdoc.unesco.org/ark:/48223/pf0000380455.

Nevertheless, since regulation of social media can determine what people can say, see and hear, legal frameworks and interventions must be well-designed and avoid overreach and negative impact on democracies. There is an urgent need for governments, social media platforms, and other businesses to protect the fundamental pillars of a democratic society and its implementing mechanisms: the elections.

The most relevant rights pertaining to electoral processes are as follows:

Universal Declaration of Human Rights

Article 21

1. Everyone has the right to take part in the government of [their] country, directly or through freely chosen representatives.

2. Everyone has the right to equal access to public service in country.

3. The will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures.

International Covenant on Civil and Political Rights

Article 25

Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2[29] and without unreasonable restrictions:

(a) To take part in the conduct of public affairs, directly or through freely chosen representatives;

(b) To vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors;

(c) To have access, on general terms of equality, to public service in [their] country.

General Comment 25 of the UN Committee on Human Rights on "The right to participate in public affairs, voting rights and the rights to equal access to public service" provide further guidance on the interpretation of article 25 of the International Covenant on Civil and Political Rights (ICCPR).

2.2. INTERNATIONAL STANDARDS AND SOFT LAW ON FREEDOM OF EXPRESSION

States' obligation to facilitate, respect and protect freedom of expression is a core component of free and fair elections. In absence of this right, electoral processes cannot proceed properly and fairly, since the right to participation cannot be fully exercised, thus undermining their validity. The right to freedom of expression is enshrined in Article 19 of the Universal Declaration of Human Rights (UDHR), which states that "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers". Moreover, Article 19 of the ICCPR – which is binding for the States that have ratified it – also protects freedom of expression, in the following terms:

[29] Article 2 ensures for all individuals under a State's responsibility the rights enshrined in the Covenant. It also establishes that this is to occur without discrimination on the stated grounds, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. The article demands domestic implementation to give full effect to those rights, with accompanying remedies for violation. See: P. Taylor, 2020, ***A Commentary on the International Covenant on Civil and Political Rights: The UN Human Rights Committee's Monitoring of ICCPR Rights***, Cambridge, Cambridge University Press, pp. 58-86.

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

General Comment No. 34 by the Human Rights Committee – which although not legally binding provides interpretative guidance of Article 19 of the ICCPR – recognizes that freedom of expression includes "all forms of audio-visual as well as electronic and internet-based modes of expression".[30] Besides, several resolutions by UN bodies have reaffirmed that "the same rights people have offline must be protected online".[31]

As noted in General Comment No. 34, freedom of opinion, which is at the core of freedom of expression, may not be neither derogated nor restricted under any circumstance.[32] However, freedom of expression is not an absolute right. It can be limited, yet very exceptionally, and considering a three-part, cumulative test.

BOX 2: THREE-PART TEST FOR RESTRICTIONS TO FREE EXPRESSION

To be in alignment with international law, restrictions to freedom of expression must:

- Be provided by law, which should be clear and accessible to everyone.
- Have a legitimate aim; that is, one of the purposes listed in Article 19, paragraph 3, of the ICCPR: to protect the rights or reputations of others; or to protect national security or public order, public health or morals.
- Be necessary and proportionate, representing the least restrictive means to achieve the purported aim.

When considering the matter of limitations to freedom of expression, attention must be paid to Article 20 of the ICCPR, which requires States to prohibit "any propaganda for war" and "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence." As noted in General Comment 34 by the Human Rights Committee, States are obliged to prohibit by law the acts referred to in this article, while also keeping strict conformity with article 19.[33]

---

[30] CCPR/C/GC/34 General Comment (GC) No. 34 on Article 19 of the ICCPR, para. 12.

[31] UN GA resolution of 27 June 2016 on the Promotion, protection and enjoyment of human rights on the Internet, A/HRC/32/L.20, para. 1, as well as; UN HRC Resolution 20.8 of 5 July 2012 and 26/13 of 26 June 2014 on the promotion and protection of human rights on the Internet, HRC Resolutions 12/6 of 2 October 2009 on freedom of opinion and expression HRC Resolution 28/16 of 24 March 2015 on the right to privacy in the digital age, GA Resolutions 68/167 of 18 December 2013 and 69/166 of 18 December 2014 on the right to privacy in the digital age and 70/184 of 22 December 2015 on the information and communications technologies for development, amongst others.

[32] CCPR/C/GC/34 General Comment (GC) No. 34 on Article 19 of the ICCPR, para. 9.

[33] Ibid., paras. 50-52.

In different reports by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the distinction is made between three types of expression: (i) speech that international law categorizes as an offence, which States are required to prohibit by law and that can be criminally prosecuted;[34] (ii) forms of expression that, while not criminally punishable, may justify limitations and a civil sanction; and (iii) expressions that do not warrant a criminal or civil punishment, but which still raise concerns in terms of tolerance, civility and respect for others. Each of these categories call for different types of legal and policy responses.[35]

Also enshrined in Article 19 of the ICCPR is the right of access to information, which is inherent to the right to freedom of expression and guarantees individuals the right to access information held by public bodies – with very limited exceptions that should be clearly and narrowly defined, subject to strict harm and public interest tests.[36] The right to information is also critical to the important democratic function of both traditional and new media, which require such access to inform public debate and provide citizens with readily available information on candidates and the electoral process. There is a direct connection between the rights to freedom of opinion and of expression, as well as access to information, and Article 21 of the UDHR and Article 25 of the ICCPR, which are the foundations of the right to participate in public affairs, to vote and to access public service.[37]

After calling attention to the role of free media as a cornerstone of democracy, the Human Rights Committee's General Comment No. 34, noted that "The free communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential. This implies a free press and other media able to comment on public issues without censorship or restraint and to inform public opinion. The public also has a corresponding right to receive media output".[38] Thus, Article 19 of the ICCPR, underpins and protects the work of journalists, including non-professional "citizen" journalists, to report freely on the conduct of election campaigns and elections, as well as any demonstrations, protest or irregularities which may occur.

Similarly, the Human Rights Committee had also stated, in its earlier General Comment No. 25 (focused on the right to participate in public affairs, to vote and to access public service) that the free flow of information and ideas between citizens, candidates and elected representatives is critical and requires media freedom, and well as individuals' freedom to debate on public issues, to criticize and oppose, to publish political content and advertise political ideas, among other preconditions.[39]

The recognition of freedom of expression and freedom of opinion as important conduits for the exercise of electoral rights, as well as for the right to associate in political parties, has also been reflected in the fact that the Human Rights Committee has considered that a violation of Article 22 of the ICCPR (freedom of association, on which the establishment of political parties is based) will also amount to a violation of Article 19 (freedom of expression) of the ICCPR, but may well also amount to a violation of Article 25 (the right to participation – including to stand for office

---

[34] I.e. "child pornography, direct and public incitement to commit genocide, advocacy of national, racial or religious hatred constituting incitement to discrimination, hostility or violence, and incitement to terrorism", https://www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf, paras. 20-36.

[35] See: https://www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf, para. 18; Report of Special Rapporteur on Freedom of Expression and Opinion, 7 September 2012 (A/67/357), https://undocs.org/en/A/67/357, para. 2.

[36] See Mendel, 2008, ***Freedom of information: A comparative legal survey*** (2 ed.), Paris, UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000158450, pp. 34-35.

[37] See p. 11 of this report.

[38] CCPR/C/GC/34 General Comment (GC) No. 34 on Article 19 of the ICCPR, para. 13.

[39] Human Rights Committee General Comment 25, ***The Right to Participate in Public Affairs, Voting Rights and the Right to Equal Access to Public Service***, CCPR/C/21/Rev.1/Add.7, 27 August 1996, para. 25.

and vote). Indeed, in periodic reports[40] the provisions of law are often dealt with together, as the freedom to associate through political parties or other organizations is seen as the foundation of participation in public life, which includes the right to stand for office or vote.

As stated before, in the context of elections in digital times, the challenges posed by disinformation bring into tension certain forms of expression and the right to vote. Free and fair elections require the free flow of information, to guarantee the electorate's free expression. However, disinformation can also undermine the right to vote. This calls for finding a balance between protecting the integrity of the right to vote while also ensuring that freedom of expression is not hampered in the process. Moreover, the development of social media and AI have also brought challenges related to privacy infringements, as well as others linked to the increased dissemination of hate speech.

2.2.1. UN RESOLUTIONS ON THE SAFETY OF JOURNALISTS

General Comment No. 34 by the Human Rights Committee defined journalism, as a "function shared by a wide range of actors, including professional full-time reporters and analysts, as well as bloggers and others who engage in forms of self-publication in print, on the internet or elsewhere".

In recent years, as the means to report expanded, journalists have also come under new dangers.[41]

The Special Rapporteurs' joint declaration in 2020 focused on freedom of expression and elections in the digital age. It expressed concerns regarding the threats and violence journalists often face during electoral periods, such as targeted smear campaigns, intimidation, and harassment offline and online, physical attacks, and called special attention to women journalists as targets.[42]

Between 2012 and 2021, multiple resolutions and decisions promoting the safety of journalists werw adopted by the UN General Assembly (2014, 2015, 2017, 2019), the UN Security Council (2014), UNESCO's governing bodies and the International Programme for the Development of Communication (2014 - 2020), and the UN Human Rights Council (2014, 2016, 2018, 2020, 2021). ***UNESCO plays a leadership role in efforts to advance the safety of journalists, including through the coordination of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity.[43] This initiative is the first systematic UN-wide plan that aims to create a free and safe environment for journalists and media workers, both in conflict and non-conflict situations, to strengthen peace, democracy, and development worldwide.***

2.2.2. THE RIGHT TO PRIVACY

The right to privacy is enshrined in Article 12 of the Universal Declaration of Human Rights, which states that "No one shall be subjected to arbitrary interference with [their] privacy, family, home or correspondence, nor to attacks upon [their] honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." In addition, this right is legally protected by Article 17 of the ICCPR, which states:

---

[40] See for example: paras. 46-49 on "Freedom of association and participation in public life" of the Concluding observations on the second periodic report of Turkmenistan CCPR/C/TKM/CO/2, 20 April 2017; and the Concluding observations of the Human Rights Committee on the Republic of Moldova CCPR/CO/75/MDA, 5 August 2002, para. 16.

[41] 2018 Report of the Special Rapporteur on violence against women, its causes and consequences on violence against women in politics, including on violence against women in elections. https://digitallibrary.un.org/record/1640483?ln=en.

[42] See: https://www.ohchr.org/Documents/Issues/Opinion/JointDeclarationDigitalAge_30April2020_EN.pdf.

[43] See: https://en.unesco.org/un-plan-action-safety-journalists.

> 1) No one shall be subjected to arbitrary or unlawful interference with [their] privacy, family, home, or correspondence, nor to unlawful attacks on [their]honour and reputation.
>
> 2) Everyone has the right to the protection of the law against such interference or attacks.

Any restriction to the right to privacy must meet the three-part test of legality, necessity and proportionality.[44] The Human Rights Committee's General Comment No. 16, elaborates on the right to privacy, reaffirming that the State has the responsibility of protecting individuals against such interferences and attacks that emanate either from the State or any other natural or legal person.[45] It also addresses data protection,[46] which is a key part of the right to privacy, although not its totality.[47] The right to privacy is seen as an enabler of the exercise of other rights, including freedom of opinion and expression, freedom of assembly and association, and political participation.

Although everyone has the right to privacy, it is generally acknowledged that politicians and other public figures should be subject under higher scrutiny. Their privacy's lower level of protection both in courts as well as vis-à-vis media coverage is justified in the interest of open discussion on matters of public concern.[48] A similar argument applies to the right to reputation – which is also enshrined in Article 17 of the ICCPR. Accordingly, UN and regional Special Mandates on Freedom of Expression, UNESCO, international and national civil society organisations (CSOs) worldwide, among others, have repeatedly called for defamation offences not to be applied in cases of criticism of public officials and, more generally, they have advocated for the full decriminalization of defamation, in favour of civil sanctions.[49] A new set of challenges to the right to privacy have arisen in the digital age, primarily using mass surveillance by States, but also through political micro-targeting and the profiling of individuals. Several Resolutions by the UN General Assembly[50] and the UN Human Rights Council,[51] as well as Reports by the UN Special Rapporteur on Freedom of Opinion and Expression[52] and the Office of the High Commissioner on Human Rights, have addressed these emerging concerns. A significant development showing this increased attention has been the creation, through a Human Rights Council resolution on the right to privacy in the digital age, of the mandate of a Special Rapporteur on the Right to Privacyin July 2015.[53]

A relevant matter concerns social messaging where encryption may protect privacy up to a point, and which may also often operate with a different logic to the algorithmic ranking of content or

---

[44] A/HRC/27/37, para. 23; Human Rights Council resolution 34/7, para. 2; Special Rapporteur Report.

[45] General Comment No. 16: The right to respect of privacy, family, home and correspondence, and protection of honour and reputation (Art. 17), adopted 4 August 1988, para. 1.

[46] Ibid., para. 10.

[47] UNHRC, A/HRC/39/29, 2018, https://undocs.org/A/HRC/39/29.

[48] UNDP, ***Media and Elections: A Guide for Electoral Practitioners***, 2015, p. 10. Available at: https://www.undp.org/content/undp/en/home/librarypage/democratic-governance/electoral_systemsandprocesses/media-and-elections--a-guide-for-electoral-practitioners.html.

[49] Civil sanctions to defamation offences should not be excessive as to provoke a chilling effect on expression, and should be proportionate to the harm caused. Moreover, non-financial remedies should be prioritized. For further reading on defamation and related international standards, see: https://www.osce.org/files/f/documents/c/b/40190.pdf; https://www.article19.org/data/files/medialibrary/38641/Defamation-Principles-(online)-.pdf.

[50] Resolution adopted by the General Assembly on 18 December 2013 [on the report of the Third Committee (A/68/456/Add.2)] para. 4. As quoted in: "The right to privacy in the digital age": Report of the Office of the United Nations High Commissioner for Human Rights.

[51] A/HRC/RES/42/15.

[52] Report of the Special Rapporteur on Freedom of Opinion and Expression, A/HRC/41/35, 28 May 2019, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/148/76/PDF/G1914876.pdf?OpenElement.

[53] See: https://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx.

recommendation engines found in open public social media. Private messaging applications and services can help civil society activists monitor and report abuses without fear. Nevertheless, they can also disseminate electoral disinformation on a big scale.

2.2.3. HATE SPEECH

There is no international legal definition of hate speech, and the characterization of what is "hateful" is controversial and disputed. According to the UN Strategy and Plan of Action on Hate Speech, the term is understood as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group based on who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor. This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive".[54]

Hate speech increases the polarization of societies. It directly affects the electoral cycles and processes and concerns EMBs, candidates, elected officials, members of civil society, political party members, activists, and voters, among other key actors.

Strategic efforts to reduce the impact of hate speech requires two important rights to be reconciled: the first is to respect freedom of opinion and expression (Article 19 of the ICCPR), which is necessary for open debate in a democratic society; and the second is the right to non-discrimination and participation in public life (Articles 2, 25 and 26 of the ICCPR).[55] The balance that needs to be struck is a difficult one.

Article 20 of the ICCPR stipulates an obligation for States to prohibit any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. Reports issued by the Office of the UN Special Rapporteur on Freedom of Opinion and Expression explain that hate speech, according to the formulation of Article 20, must include three key elements: 1) advocacy of hatred, 2) advocacy which constitutes incitement, 3) incitement that is likely to result in discrimination, hostility, or violence.[56] Yet balancing freedom of expression and the prohibition of incitement to hatred is no simple task.[57]

Considering the complexities involved in interpreting what constitutes hate speech and in identifying how to best tackle it, the OHCHR organized several multi-stakeholder workshops that resulted in the Rabat Plan of Action on the prohibition of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence.[58] It also contains threshold tests and recommendations, which are extremely relevant to social media and many other aspects of the digital universe.

A high threshold emerged from this process for defining restrictions on the freedom of expression, incitement to hatred, and how to apply Article 20 of the ICCPR; which consists of a six-part test. It takes into consideration: the context, the speaker, the intent, the content and form of the speech,

---

[54] See: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml.

[55] Report of Special Rapporteur on Freedom of Expression and Opinion, 7 September 2012 (A/67/357), https ://undocs.org/en/A/67/357; Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, on hate speech online, 9 October 2019 (A/74/486), https ://undocs.org/A/74/486.

[56] See: https://undocs.org/en/A/67/357, para. 43; https://undocs.org/A/74/486, para.8.

[57] UNESCO, 2015, ***Countering Online Hate Speech***, p. 20. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000233231.

[58] Freedom of expression vs incitement to hatred: OHCHR and the Rabat Plan of Action. Available at: https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action. See also: https://www.youtube.com/watch?v=ADrB32O-Se3A&t=4s.

the extent of its dissemination, and the likelihood that it could incite harm, including imminence.[59] In addition, restrictions to freedom of expression in cases involving incitement to hatred should also still meet the conditions of legality, necessity and proportionality, and legitimacy.[60]

While Article 20 of the ICCPR obligates to prohibit hate speech by law, it does not obligate States to criminalize it.[61] In this regard, the Rabat Plan of Action clarifies that criminal sanctions related to unlawful forms of expression should be seen as the last resort measures to be applied only in strictly justifiable situations. Civil sanctions and remedies should also be considered, including pecuniary and non-pecuniary damages, along with the right of correction and the right of reply. Administrative sanctions and remedies could also be implemented, including those identified and put in force by various professional and regulatory bodies.[62] Similar recommendations have been made in reports by the UN Special Rapporteur on Freedom of Opinion and Expression.[63]

In line with the above comprehensive approaches, in June 2019 the UN Secretary-General launched the UN Strategy and Plan of Action on Hate Speech. It outlines the range of actions through which the UN can contribute to preventing and countering hate speech, all while protecting the right of freedom of opinion and expression.[64]

There are other international treaties, besides the ICCPR, with relevant provisions for defining hate speech and identifying responses to it:

- The Convention on the Prevention and Punishment of the Crime of Genocide
- The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)
- The Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)
- The Convention on the Rights of Persons with Disabilities (CRPD)

While a legal response is essential, legislation is only part of a more extensive toolbox to respond to the challenges of hate speech. Any related legislation should be complemented by initiatives from various sectors of society geared towards a plurality of policies, practices, and measures nurturing social consciousness, tolerance and understanding change, and public discussion. This is with a view to creating and strengthening a culture of peace, tolerance and mutual respect among individuals, public officials and members of the judiciary, as well as rendering media organizations and religious/community leaders more ethically aware and socially responsible. States, EMBs, political parties and groups, media and society have a collective responsibility to ensure that acts of incitement to hatred are spoken out against and acted upon with the appropriate measures, in accordance with international human rights law.

---

[59] See: https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf, para. 29.

[60] See: https://undocs.org/en/A/67/357, para. 41; https://undocs.org/A/74/486, para. 6; https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf, paras. 18 and 22.

[61] See: https://undocs.org/en/A/67/357, para. 47; https://undocs.org/A/74/486, paras.15-18.

[62] See: https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf, para. 34.

[63] See: https://undocs.org/en/A/67/357; https://undocs.org/A/74/486; https://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf.

[64] Ibid., pp. 3-5.

BOX 3: THE ADDIS ABABA DECLARATION ON "JOURNALISM AND ELECTIONS IN TIMES OF DISINFORMATION"[65]

An outcome of the 2019 World Press Freedom Day commemoration,[66] the Addis Ababa Declaration on "Journalism and Elections in Times of Disinformation" emerged from discussions held by over 2,000 participants from 100 countries who were convened at this international event.

The Declaration calls attention to the fundamental role that free, independent and pluralistic journalism – exercised both online and offline – plays in democracy. It acknowledges the potential benefits that the development of ICTs signifies for electoral processes while also noting the growing risks related to hate speech, disinformation and data collection, as well as the potential risks imposed by the utilization of social media and social messaging to undermining citizen's ability to make informed decisions, thereby compromising electoral fairness.

It calls on UNESCO Member States to dismantle legal obstacles to freedom of expression, and to avoid regulatory efforts that are worded in a vague manner as well as measures that are not aligned with the principles of legality, necessity and legitimacy, and proportionality. It also calls on them to avoid delegating to Internet Service Providers the regulation of online content in a way that does not abide by international human rights law.

Referring to the dangers faced by journalists, press cartoonists, artists, activists and other actors who publicly express themselves, which is of particular relevance in connection to elections, it calls for the establishment of transparent and effective systems to be established for their protection.

The Declaration warns about the potential negative impact that an over-regulation of digital electoral communications can have on freedom of expression and privacy rights and highlights the importance of promoting Media and Information Literacy among citizens.

The document calls for the development of guidelines and policies for media and Internet companies' use of AI tools, given the implications that these may have on human rights. It emphasizes the importance of online content verification via independent journalism and further calls on journalists, media outlets, electoral practitioners, and Internet intermediaries to expose disinformation and propaganda, tackle the issues posed by filter-bubbles, take steps to bring greater transparency into political advertising as well as into Internet companies' terms of service and other policies.

---

[65] See: UNESCO Addis Ababa Declaration World Press Freedom Day 2019, "Journalism and Elections in Times of Disinformation", UNESCO World Press Freedom Day International Conference, held in Addis Ababa, Ethiopia, 1-3 May 2019, https://en.unesco.org/sites/default/files/wpfdaddisdecl3_may.pdf.

[66] See: https://en.unesco.org/sites/default/files/wpfd2019_concept_note_en.pdf, https://en.unesco.org/events/world-press-freedom-day-2019.

2.3. AI CHALLENGES FOR HUMAN RIGHTS AND ELECTIONS

Although international human rights legal frameworks were developed with a focus on States, there has been increasing attention to enterprises and businesses' human rights responsibilities over the past decades.

Several initiatives have therefore emerged to guide companies in this regard. Among these are the UN Guiding Principles on Business and Human Rights, which the Human Rights Council endorsed in 2011. According to these principles, the primary duty to protect human rights lies with States. Yet, businesses have a responsibility to respect these rights, and both States and companies have a role in facilitating access to remedy for those whose rights are violated. The principles reaffirm that governments must guarantee that not only state organs respect human rights but that companies operating under their territory/jurisdiction do so as well. These principles are also of relevance when considering the role and responsibilities of Internet intermediaries like social media companies.[67]

In this context, elections face new digitally driven dangers, including online attacks on and harassment against journalists, candidates, activists, voters, etc., aimed at deterring them from freely and impartially participating during elections or undermining EMBs' reputation and electoral legitimacy as a whole. The combination of hate speech and disinformation can trigger tensions resulting in electoral-related conflict and violence. Accordingly, there have been increasing demands for enhanced algorithmic accountability.

It is also true that AI has great potential for enhancing independent journalism, campaigning, and supporting electoral processes in general.[68] Algorithms have a positive impact when used to reduce the visibility or remove content that discriminates or incites hate and violence. However, the use of AI might entail the risk of blocking legitimate forms of expression, limiting the circulation of legitimate content, democratic debate, and pluralism during electoral periods as algorithms cannot fully assess all content, such as detecting all semantic nuances of communication (e.g., ironic remarks, jokes, etc.).[69] Algorithms and the manner in which they are developed by private companies should be more transparent and accessible in order to disperse any doubts of bias

The right to political participation not only requires freedom of expression, as stated by the UN Human Rights Committee's General Comment No. 25, but also presupposes that "persons entitled to vote must be free to vote for any candidate for election and for or against any proposal submitted to referendum or plebiscite, and free to support or to oppose the government, without undue influence or coercion of any kind which may distort or inhibit the free expression of the elector's will. Voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind".[70] Interference and manipulation through disinformation disseminated via social networks and powered through AI during electoral processes greatly impact on the right to vote freely.

The potential of automated decision-making systems to reinforce bias and discrimination also impacts on the right to equality and the right to participation in public life,[71] for example, to "disproportionately harm historically underrepresented communities".[72] AI technology also has important gendered implications which, coupled with the existing gender divides in terms of

---

[67] OHCHR, 2011, ***UN Guiding Principles on Business and Human Rights***. Available at: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

[68] X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019, ***Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective***, UNESCO.

[69] See: https://undocs.org/A/73/348.

[70] Human Rights Council General Comment 25, para. 19.

[71] See: https://undocs.org/A/73/348; X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019.

[72] See: https://undocs.org/A/74/486, 2019.

digital access and digital skills, as well as the specific online risks faced by candidates, politicians, activists, public figures and journalists who identify as women, have an impact on women's rights and gender equality overall.

The right to privacy has come under new threats with the emergence of AI-driven methods for data collection, de-anonymization, tracking, and profiling, in a context where algorithmic transparency is lacking, and the effectiveness of data protection regimes varies worldwide.[73]

Each individual should have the possibility of seeking recourse to an independent tribunal or court in instances of a violation of their rights. Any internal mechanism to appeal the take-down or blocking of content should not hamper an individual's right to seek redress to a possible violation of rights through a domestic or international or regional court, as appropriate.

AI poses problems regarding the right to a remedy. Internet users are most often unaware of how algorithms impact the information they access via social media. The logic underlying an algorithmic decision is challenging to understand even for experts, as companies update algorithms frequently, and AI applications may also modify their algorithms, all without transparency. In addition, there is a move towards the use of automated remedy systems to handle complaints by users, which causes further concerns, as these mechanisms do not possess the discretion nor the capacity to analyse context and take independent decisions.[74]

Remedies involving restrictions on free speech and on political and electoral rights might be controversial, as they may limit fundamental rights in a democratic society. When it comes to the normative framework that applies to social media and AI in elections, the analysis should consider the international standards enshrined in national, regional, and international frameworks, rules and regulations, and the terms of service and community standards that guide social media companies' self-regulation.[75]

2.4. CONCERNS REGARDING DEMOCRATIC ACCOUNTABILITY

Personal data drives much of the Internet economy, where the services of social platforms, search engines and other Internet intermediaries are offered for free in exchange for the use of information that consumers provide about themselves, in many cases unknowingly or unwillingly.[76] Internet platforms thus monetize the personal data of a large number of customers, using it for marketing purposes, and increasingly also for political ones, such as political micro-targeting. Monopolies are established as the so-called networking effects enhance the effectiveness of the provided services thanks to having a higher number of data sets, which in turn results in an even higher number of users that lead to an increase in available data sets – feeding into the cycle once again. Political micro-targeting fuelled by AI, driven by aggregated personal data that is not always collected in lawful ways and coupled with limited users' protection, gives increasing power to big technology companies and governments to track people's conduct, views and contacts online.[77] In this context, tensions between freedom of expression and the protection of the rights of others – including reputation, privacy, data protection, and intellectual property rights – have also increased.[78]

---

[73] X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019.

[74] See: https://undocs.org/A/73/348.

[75] See: https://democracy-reporting.org/en/office/global/collection?type=.

[76] UNESCO, World Trends in Freedom of Expression and Media Development, Global Report 2017/2018. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065.

[77] UNESCO: World Trends in Freedom of Expression and Media Development, Global Report 2017/2018. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065; T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

[78] https://compact-media.eu/wp-content/uploads/2021/02/D2.2.pdf.

This model has a significant impact on electoral processes, as the profiling and micro-targeting of voters – which was analysed earlier in this Guide – affects their capacity to make informed decisions.[79] It may exclude a person from the "marketplace of ideas" and result in voter bias due to the more targeted or limited exposure to information, as well as fragment political debate and allow political parties to send contradictory messages to voters. Likewise, voters' personal data feeds into strategic analysis and action without their formal consent, which has raised concerns regarding their vulnerability and the lack of transparency in relation to how a voter's data was obtained and why they are being targeted. Additionally, AI's capacity to identify patterns and trends can be accurate to the point of de-anonymizing users or groups of people whose information is continuously tracked.[80] The deployment of data-driven methods in election campaigns, the increasing integration of social media platform companies, and their great influence on political beliefs and behaviours, raise important concerns about democratic accountability.[81] Furthermore, micro-targeting undermines the impact of regulating advertising on broadcast media to enhance its fairness and transparency. There have been increasing calls for political micro-targeting, like other forms of political advertising, to be subject to official campaign financing legislation. The issue is, however, that regulation of political micro-targeting might get in conflict with the right to freedom of expression of political opinion.[82]

The growing attention to these matters has been reflected in important statements, decisions, declarations, and initiatives by international bodies. In 2017, the Human Rights Council emphasized that:

"[the] unlawful or arbitrary surveillance and/or interception of communications, as well as the unlawful or arbitrary collection of personal data, as highly intrusive acts, violate the right to privacy, can interfere with other human rights, including the right to freedom of expression and to hold opinions without interference, and the right to freedom of peaceful assembly and association, and may contradict the tenets of a democratic society, including when undertaken extraterritorially or on a mass scale."[83]

In turn, the Council of Europe's Committee of Ministers issued a Declaration, on 13 February 2019, on the manipulative capabilities of algorithmic processes, warning that:

> "Contemporary machine learning tools have the growing capacity not only to predict choices but also to influence emotions and thoughts and alter an anticipated course of action, sometimes subliminally. The dangers for democratic societies that emanate from the possibility to employ such capacity to manipulate and control not only economic choices but also social and political behaviours, have only recently become apparent."[84]

According to the Committee, this calls for democratic oversight of such systems, particularly in the context of elections, in order to safeguard their fairness and integrity by ensuring voters' access to comparable levels of information and their protection against manipulation and other unfair practices.[85] The Council of Europe's Venice Commission has also warned that the current configuration of social media platforms "allows for political advertising to be increasingly individually tailored and targeted. Instead of being a public square featuring many voices, people are becoming more isolated and out of touch with the whole spectrum of the public."[86]

---

79 T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***.

80 X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019.

81 C. J. Bennett and D. Lyon, 2019, Data-driven elections, ***Internet Policy Review***, Vol. 8, No. 4.

82 OSCE/ODIHR, 2015.

83 See: https://www.un.org/ga/search/view_doc.asp?symbol=A/HRC/34/L.7/Rev.1, p. 3.

84 Council of Europe the Declaration by the Committee of Ministers on "the manipulative capabilities of algorithmic processes", 13 February 2019, Decl(13/02/2019)1, para. 8.

85 Ibid.

86 EC (Venice Commission), 2019, ***Joint Report of the Venice Commission and of the Directorate of Information Society and Action against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital***

In April 2016, the European Parliament and the Council of the European Union adopted the General Data Protection Regulation (GDPR), which sets a high standard within the EU and could be an example to push for more transparency of online political advertising. The GDPR enshrined the notion of informed consent, whereby all people should have more control over what personal data about them is being collected and why, as well as to be able to correct or erase data that is being held about them.[87] While the GDPR's enforcement is delegated to national data protection authorities and they have different levels of experience and expertise, a regional body, in this case the EU, must ensure consistent implementation by providing further guidance to its member States. Since the GDPR was passed, enforcement measures have been put into practice to ensure that data infringements are punished.[88]

Although the GDPR presents the right step forward and limits the purchase of personal data, experts have pointed out that, in certain aspects, it is not sufficient and that further regulations would be necessary to ensure protection for personal data.[89] No other regional private data protection regulation comparable to the GDPR exists, and countries regulate this area nationally. South Africa's Protection of Personal Information Act (POIPA) is an example of regulation that has been modelled after the GDPR.[90]

ACTIVITY II

The following activity has the objective of determining the participant's level of knowledge on the international human rights law framework, the new challenges to human rights in the digital era, the relevance of women's rights and political participation, freedom of expression, safety of journalists, hate speech, and main concerns about democratic accountability.

Suggested guiding questions for a discussion:

I. How do you observe the impact of AI on human rights in the context of electoral processes in your country?
II. Can the right to privacy be affected during elections by AI and social media? Please provide examples.
III. Why is women's political participation capital for democracies and how can it be affected by social media? Is there a risk for it to be disproportionally affected and if so, why? You can also take into account intersectional perspectives.
IV. The Addis Ababa Declaration about journalism and elections in times of disinformation calls attention to the fundamental role that free, independent and pluralistic journalism plays in democracy.
V. Can you explain the positive and negative impact of regulations on digital communications?
VI. Which are the main concerns regarding democratic accountability? Please provide examples based on your own experience.

Technologies and Elections, CDL-AD (2019)016, para. 14.

87 See: A. Puddephatt, 2019, ***Social media and elections***, UNESCO, Paris. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000370634.

88 T. Dobber, R. Ó Fathaigh and F. J. Zuiderveen Borgesius, The regulation of online political micro-targeting in Europe, ***Internet Policy Review***, 2019.

89 C. J. Bennett and D. Lyon, 2019.

90 B. McKenzie, ***General Data Protection Regulation (GDPR) in Africa: So What?***, 2019. Available at: https://www.lexology.com/library/detail.aspx?g=f9d05505-ae8c-473e-a322-40c376fd8217.

3. THE NEW INFORMATION PARADIGM LAW FRAMEWORK

OBJECTIVES OF THIS SECTION

- Identify the main characteristics of the new information paradigm, the digital divide and the gender divide.
- Understand the concepts disinformation, misinformation and mal-information and the drivers behind misleading content.
- Examine the actors behind and the targets of disinformation during and within the electoral cycle.

The popular term "fake news" has been criticized for being too broad and vague, which leaves it susceptible to arbitrary use and misuse.[91] Politicians and other actors sometimes abuse this terminology as an accusation designed to undermine the reputation and credibility of individual journalists, individual media organizations and/or the particular information at stake.[92] The manipulation of the rhetoric of the "fake news" has been reframed as part of a more comprehensive approach that considers "fake news" as disinformation and as one element of the digital era and the new information paradigm.[93]

The spread of disinformation – which has been exacerbated via its circulation through social media platforms, powered by AI – has become a critical challenge. It casts a shadow over the integrity of public debate and elections, undermining citizens' trust in democratic institutions and in the media, while also negatively impacting the accuracy and reliability of the information that feeds public opinion. It has the potential to deepen existing societal and political polarization, as well as to generate confusion among voters, challenge fact-based information and undermine candidates, institutions, and vulnerable groups.[94]

The massive volume and reach of disinformation and misinformation dressed up as news and distributed via social media has inflicted further reputational damage to journalism and undermined democracies worldwide. In the high-speed information free-for-all on social media platforms and the Internet, everyone can be a publisher. As a result, citizens struggle to discern what is true and what is false. Extreme views, conspiracy theories and populism flourish, and once-accepted truths and institutions are questioned. In this world, newsrooms battle to claim and perform their historic role as gatekeepers whose products can help to establish the truth. At the same time, the rise of marketplaces for "strategic communications" and "information operations", including active disinformation and mal-information, has become a major factor in the information ecosystem.[95]

3.1. DISCERNING DIFFERENCES

While the historical impact of rumours and fabricated content has been well documented, new technologies have attained an effect never seen before on content pollution on a global scale.

---

[91] J. Posetti, C. Ireton, C. Wardle, H. Derakhshan, A. Matthews, M. Abu-Fadil, T. Trewinnard, F. Bell and A. Mantzarlis, Journalism, ***'Fake News' & Disinformation - Handbook for Journalism Education and Training***, UNESCO, 2018. Available at: https://en.unesco.org/fightfakenews.

[92] Edelman Trust Barometer - Global Results, (online), 2017. Available at: https://www.edelman.com/global-results/ (accessed 03/04/2018).

[93] Ibid.

[94] C. Wardle and H. Derakhshan, ***Information Disorder: Toward an interdisciplinary framework for research and policymaking***, Council of Europe, 2017. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[95] J. Posetti et al., 2018. Available at: https://en.unesco.org/fightfakenews.

Developments in the last few years have placed journalism under fire.

A range of factors are transforming the communications landscape, raising questions about the quality, impact and credibility of journalism. At the same time, orchestrated campaigns are spreading untruths - disinformation, mal-information and misinformation - that are often unwittingly shared on social media:

- **Disinformation:** Information that is false and deliberately created to harm a person, social group, organization, or country.[96] The motivations underlying it could be to make financial profit, to have foreign or domestic political influence, or simply to cause trouble.[97]
- **Misinformation:** Information that is false, but not created with the intention of causing harm.[98] Often, disinformation turns into misinformation, when individuals share false or misleading content that they believe is correct without realizing that it is inaccurate or misleading. This sharing tends to be motivated by social and psychological factors, as people tend to "perform" their identities online (e.g., through likes, comments or shares) and want to feel connected to their "group" or a certain community, for example, groups of people sharing the same ideas or concerns about certain topics, who are part of the same political party, or of the same religion, race or ethnic group.[99]
- **Mal-information:** Information that is based on reality, used to inflict harm on a person, organization, or country.[100] It often aims to cause harm by making public content that was designed to stay private.[101]

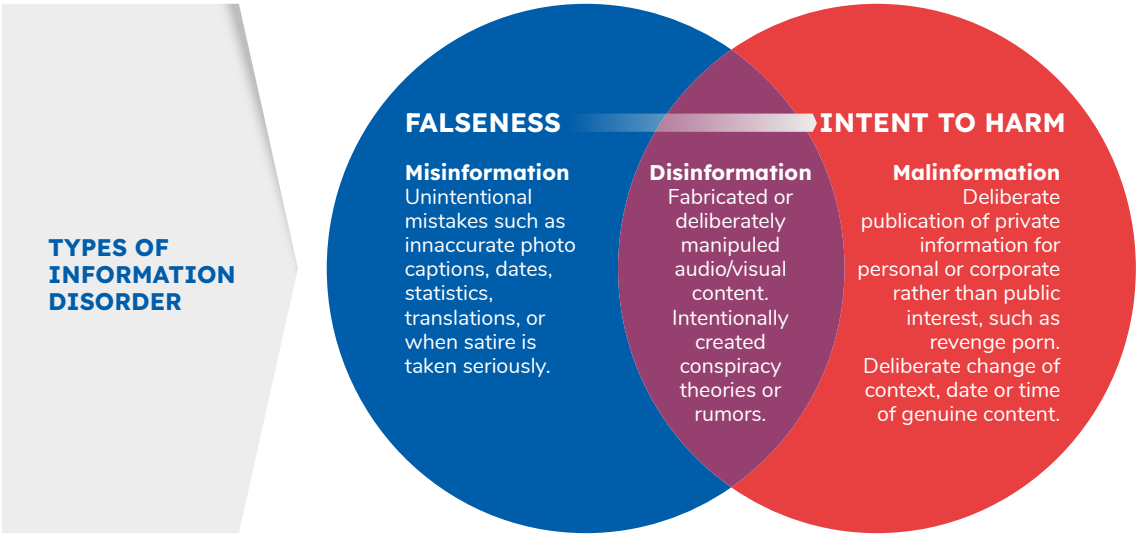


FIGURE 3: TYPES OF FALSE AND MISLEADING CONTENT[102]

---

[96] See C. Wardle and H. Derakhshan, 2017. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c, p. 20.

[97] See: https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x19182, p. 8.

[98] Ibid.

[99] See: https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf?x19182, p. 8.

[100] Ibid.

[101] Ibid.

[102] Ibid.

Further, this conceptual framework examines three key elements:

- **The Agent:** This first element considers the actors that create, produce, and distribute false and misleading content, as well as their motivations, intended audiences, and the use of automation.
- **The Message:** The content that is being spread, its durability, accuracy, legality, whether it involves posing as an official source, and its intended targets.
- **The Interpreter:** this element focuses on how those who receive the message make sense of it, and the actions they take in relation to it, if any.

FIGURE 4: ELEMENTS OF FALSE AND MISLEADING CONTENT[103]

Also, according to First Draft News framework, this analysis needs to consider three phases to produce misleading content: creation, production, and distribution.[104]

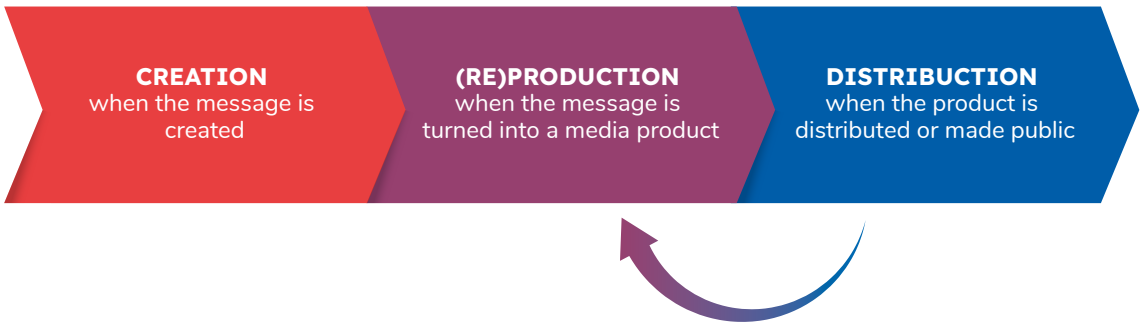


FIGURE 5: PHASES OF THE CREATION OF MISLEADING CONTENT[105]

[103] C. Wardle and H. Derakhshan, 2017. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[104] Ibid.

[105] Ibid.

The Broadband Commission research report on countering digital disinformation while respecting freedom of expression[106] produced by UNESCO, builds on First Draft News' conceptualization, as well as on the "ABC" framework[107] that distinguishes between Actors, Behaviours and Content, adapting and combining these models as well as integrating two additional elements:

- **Instigators:** actors who are at the origin of the creation and distribution of disinformation, and benefit from it, and who often pay for it to be operationalized. Sometimes they are the same as the agents, but this is not always the case, as the latter may be individuals who are paid or contracted, or who offer voluntary support or participate unintentionally in creating and spreading the content.
- **Intermediaries:** vehicles for the content, such as social media platforms or applications. Aspects to be considered under this element include the extent to which disinformation is jumping across intermediaries, how it is spreading, algorithmic and policy features that are being exploited, if there is evidence of coordinated behaviour, whether responses limit free speech, intermediaries' transparency, and accountability, among others.

This approach analyses the dynamic between Instigators, Agents, Messages, Intermediaries, Targets/Interpreters, to shed further light on disinformation's "complete lifecycle – from instigation and creation to the means of propagation to real-life impact", as well as on the responses that seek to counter it.[108]

3.2. DRIVERS BEHIND MISLEADING CONTENT

Social media platforms, AI and social messaging have changed how information is produced, communicated, and distributed.[109] In this context, despite not being new phenomena, disinformation and misinformation have grown exponentially.

Users curate their own content streams - including content from news services, journalists and other reliable information providers - without mediation, although within overdetermining algorithmic parameters. As a result of distribution via 'trust networks' (users and peers), inaccurate, false, malicious and propagandistic content masquerading as news has found increased traction. Researchers have discovered that both emotive content, and content shared by a friend or family member, is more likely to be redistributed on social media.[110] Unfortunately, these increase the likelihood of disinformation and misinformation going viral with distribution amplified by 'trust networks'[111] and emotional reactions (e.g., triggered by confirmation bias) contributing to the new communication paradigms. Among important drivers of the situation are the following:[112]

---

[106] K. Bontcheva and J. Posetti (eds.), 2020, ***Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression. Broadband Commission research report in 'Freedom of Expression and Addressing Disinformation on the Internet***, pp. 27-29. Available at: https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf.

[107] See: https://t.co/6Lb7DROYQf.

[108] Ibid., p. 27.

[109] D. Gillmor, 2004, ***We, the Media: Grassroots Journalism By the People, For the People***, O'Reilly, http://www.authorama.com/we-the-media-8.html.

[110] V. Bakir and A. McStay, ***Fake News and the Economy of Emotions Digital Journalism***, Taylor and Francis, July, 2017, http://www.tandfonline.com/doi/abs/10.1080/21670811.2017.1345645.

[111] 'Trust networks' are networks of people sharing information online via trust-based relationships (e.g. family and friendship groups) in an unmediated manner, peer-to-peer. Research has repeatedly demonstrated that social media users are more likely to share information derived from such 'trust networks' regardless of whether it's accurate or verified.

[112] This section draws from C. Wardle and H. Derakhshan, 2017 (available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c) and B. Martin-Rozumilowicz and R. Kužel, ***Social Media, Disinformation and Electoral Integrity***, IFES working paper, 2019, pp. 10-11. Available at: IFES, Working Paper, 'Social Media, Disinformation and Electoral Integrity', August 2019.pdf.

- It is now easier to create and distribute content, given the increased availability of technologies for editing and publishing.
- As opposed to traditional media, on social media platforms the content does not go through an editorial overview guided by ethical standards that require journalism to seek the truth, distinguish fact from opinion and ensure fact-checking, among other aspects.
- The fast-tracked news cycle and the use of mobile phones has accelerated the dissemination of information. Content is shared in real-time between friends, peers, family members, etc.
- On social media platforms, people's information consumption have gone from being private to public. Consequently, there is a "performative" aspect to behaviour on these platforms, like sharing, liking, and commenting on certain posts affects how a person is perceived by the members of their networks. This is compounded by a human tendency to want to conform, belong to a certain group, as well as by a "confirmation bias" – that is, the preference to consume information sources that support prior beliefs.
- Humans prefer to connect with those who share perspectives that are like theirs. They thus choose to spend more time in "echo chambers", which they perceive as a safe space to express themselves and where they face fewer conflicting views. Within echo chambers, the challenging of ideas is rare.
- In turn, social media companies benefit the most when users remain connected to their platforms, as this maximises users' exposure to ads. Thus, social media companies use algorithms to show users content of the type they have already shown to prefer via their likes, comments, or shares. These AI-driven techniques, often referred to as "filter bubbles", tend to keep social media users in their "echo-chambers", consuming content that predominantly validates their views.[113]
- The emergence of a "click-bait" online business model also serves to fuel disinformation, since sensationalist headlines tend to attract users to a wider degree, including to content that is false or misleading.
- The disruptions suffered by the traditional business models of numerous media outlets are also a relevant factor, given declining revenues that frequently shift towards the online sphere. This tendency negatively impacts quality journalism and makes producing click-bait content more tempting, further weakening the defences against disinformation.
- Much of the disinformation circulates in visual content (photos, videos, memes, etc.) and can be particularly persuasive. Visual content is more often shared and is favored over text by social platforms' algorithms. In addition, humans process visual content faster, often leading to emotional responses which diminishes the probability for them to use analytical skills. Furthermore, it can be up to impossible to trace the source of, for example, viral memes.
- Bots contribute to making content viral and target users. They can manipulate algorithms to give certain information an aspect of popularity or an impression of a widely shared belief around a specific matter, which users are likely to disseminate further.

---

[113] There is sometimes confusion between the notions of "echo chambers" and "filter bubbles", which, although closely linked, are different and non-interchangeable. "Filter bubbles" make reference to how algorithms limit the variety and range of information that a person sees for instance in social networks, given that AI-powered predictions of a user's preferences prioritize certain content and exclude other. This restricts a person's exposure to certain information and views, therefore naturally contributing to the emergence of "echo chambers" because they can serve to strengthen a person's prior opinions by over-exposing them to a certain kind of political content, reaffirming already held perspectives. X. Hu, B. Neupane, L. Echaiz, P. Sibal and M. Rivera Lam, 2019, ***Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective***, UNESCO, p. 33.

- Another worrying trend is what has been called a "weaponization of content": information reframed in misleading ways, which makes it less prone to be identified by AI systems as disinformation.
- Another concerning trend is the increased use of social messaging apps like WhatsApp, Telegram, or Signal to spread disinformation, hate speech, and polarizing narratives to fuel tensions. As social messaging apps facilitate encrypted private communications, disinformation channelled through them is hard to detect, and it is challenging to take action against those behind it. At the same time, encryption is essential to protect privacy and freedom.[114]

Even though social media platforms have become dominant vectors of disinformation, mainstream media plays a significant part in amplifying it, either unwittingly or intentionally.[115]

3.3. ACTORS THAT INSTIGATE, PRODUCE, AND SPREAD DISINFORMATION

The instigators of disinformation and the agents serving to operationalize it can be national and internal, or foreign and acting from abroad. It may involve foreign governments, criminal groups, politicians, political parties, government officials, influencers, news organizations, disinformation websites, Public Relations firms, interest groups, conspiracy theorists, violent extremists' groups, individuals, and others (e.g., political rivals and geopolitical adversaries) seeking to contribute to a country's destabilization, undermine democracy, fuel violence or generate a certain election result. Furthermore, trolls (that is, "human-controlled accounts performing bot-like activities" or harassing others online) have also emerged as crucial agents spreading disinformation.[116]

The incentives for internal actors to instigate or spread disinformation can be political (e.g., to undermine opponents and push an agenda), economic (to make a profit), or issue-based (to serve ideological or other goals – such as testing the system). When it comes to the motivations of external actors, a foreign State may, for example, try to infiltrate geopolitical adversaries' election systems (hard technology) and influence public opinion through relentless use of disinformation across social media (soft technology), all along the electoral cycle. Other, non-state, external actors may be motivated to spread disinformation across borders to promote violent extremism, as well as to make money (e.g., by creating websites supporting a candidate in elections abroad and filling them with sensationalist and deceptive content).

The complexity is that under international law, much disinformation is not illegal, except where it incites violence, hatred, or discrimination; violates a right such as right to reputation; or threatens public order. Indeed, robust political contestation often involves political actors taking information out of context and resorting to exaggeration. Responses to electoral disinformation should not intrude on legitimate competing political narratives and debates.

However, misleading content and cumulative falsehoods - such as claims about an election being stolen - can build momentum that over time can serve as "dog-whistles" for violent insurrection, in direct contrast to elections being modalities for peaceful conflict resolution. As has been argued, the Internet companies themselves sit on the data which can show when unsubstantiated political claims can convert into risks to elections as such. This raises questions about what they do with such information and whether EMBs or others can or should be able to access it.

---

[114] Ibid. and J. Posetti, 2017, ***Protecting journalism sources in the digital age***, UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000248054.

[115] C. Wardle and H. Derakhshan, 2017. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[116] Ibid.

3.4. TACTICS AND TECHNIQUES TO SPREAD DISINFORMATION

Approaches that are often used to spread disinformation, including in electoral processes:[117]

- **Coordinated inauthentic behaviour** (CIB) has been among the methods exploited for political gain by different actors in recent years, and can target domestic audiences in their own countries, or audiences in another country. It takes place when actors coordinate among themselves and use fake accounts as a key element within their operations to mislead people regarding who they are and what they are doing.[118] Internet trolls, who can individually harass, provoke or intimidate others – often to generate distraction or discord – often also engage in CIB.[119] For example, instigators of disinformation deploy "troll farms" or "troll factories" (coordinated groups of trolls disseminating certain narratives) to get content trending, generate online debate,[120] influence public opinion and decision-making.[121]
- **Information operations** entail "the collection of tactical information about an adversary as well as the dissemination of propaganda in pursuit of a competitive advantage over an opponent".[122] Social media platforms have adopted this concept to refer to "organized communicative activities that attempt to circulate problematically inaccurate or deceptive information on their platforms".[123] A similar concept is that of influence campaigns, which are "actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome".[124]
- The **computational amplification of disinformation** with bots or fake accounts to share and promote content contributes to the problems caused by the above-mentioned tactics. For example, it could serve to make some extreme views that would normally be filtered by traditional media's editorial boards part of the mainstream debate on social media. Distinct from these exploitations of platforms, the algorithms by which Internet companies curate their services, can give priority and visibility to disinformation, and downplay informational content such as verified news and informed comments.
- **Clickbait** is marketing, advertising or information material that generates interest and, thus, engagement, through a sensationalist headline that attracts clicks.
- **Astroturfing** is another method for online manipulation, which has been defined as an "organized activity that is intended to create a false impression of a widespread, spontaneously arising, grassroots movement in support of or in opposition to something (such as a political policy) but that is in reality initiated and controlled by a concealed group or organization."[125] State actors have used astroturfing campaigns that deploy troll factories, click farms (companies that employ a group of people to click on certain content, create fake profiles and posts in order to promote certain narratives) and automated social media accounts.[126]

---

[117] See also K. Starbird, A. Arif and T. Wilson, 2019, ***Disinformation As Collaborative Work: Surfacing The Participatory Nature Of Strategic Information Operations***.

[118] See: https://about.fb.com/news/2020/10/removing-coordinated-inauthentic-behavior-september-report/.

[119] B. Martin-Rozumilowicz and R. Kužel, 2019, pp. 10-11. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.

[120] Ibid.

[121] Ibid.

[122] See: https://www.rand.org/topics/information-operations.html.

[123] K. Starbird, A. Arif and T. Wilson 2019, p. 2.

[124] C. Wardle and H. Derakhshan, 2017, p. 16. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[125] See: https://www.merriam-webster.com/dictionary/astroturfing.

[126] S. Bradshaw and P. Howard, ***Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media***

- **Manipulation of search rankings** on search engines or social media platforms, which aims to make content more likely to be shown, through the manipulation of algorithms.[127]
- **Selective censorship** seeks to ensure that certain topics are left out of online conversations, by removing specific content from a platform.
- **Hacking and sharing damaging information**,[128] as well as **doxing** – which entails publishing personal information about people without their consent, which can include addresses, phone numbers, credit card details, medical information, private e-mails, etc.. As this consists of accurate information, which is disseminated publicly to inflict harm, doxing is an example of mal-information.[129]
- The use of **hashtags** and the creation of "**mutual admiration societies**" (groups of websites linking to each other, and accounts that follow and share each other's content) serves to promote certain messages.[130]
- **Micro-targeting** of advertising messages to persuade, mobilise them, impact their political opinions and voting.
- **Impersonation of fact-checking organizations, media outlets, individuals and governments** through false websites and social media accounts, as well as through bots.[131]
- **Voter suppression** is a method used to impact an election's results by preventing or discouraging certain groups from voting and can be operated – among other ways – through the dissemination of content through social media or social messaging apps (e.g., by sharing incorrect information about the location, time of voting, etc.)

3.4.1. TYPOLOGY OF DISSEMINATED CONTENT

False claims and textual narratives often (but not always) mix strong emotional language, lies and/or incomplete information, and personal opinions, along with elements of truth. It is especially difficult to uncover this modality of disinformation when it circulates on closed social messaging apps.[132] There are different formats of disinformation,[133] considering the content's modality as well as the way it is created or manipulated:

- **False or misleading narratives that aim to pass like news articles or documentary content.** Deliberately publish misleading, deceptive, or incorrect information purporting to be real news about politics, economics or culture. This content includes ideologically extreme, hyper-partisan, or conspiratorial news and information, as well as various forms of propaganda. In order to

Manipulation, Oxford University, August 2017. Available at: https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf.

[127] B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: IFES, Working Paper, 'Social Media, Disinformation and Electoral Integrity', August 2019.pdf; J. Tucker, A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal and B. Nyhan, 2018, ***Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature***.

[128] B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: IFES, Working Paper, 'Social Media, Disinformation and Electoral Integrity', August 2019.pdf.

[129] C. Wardle, 2018, ***Information Disorder: The essential glossary***. Available at: https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf.

[130] J. Tucker, A. Guess, P. Barberá, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal and B. Nyhan, 2018.

[131] K. Bontcheva and J. Posetti (eds.), 2020.

[132] Ibid., p. 22.

[133] Ibid.

catch attention, they are very often accompanied by moving images, excessive capitalization, personal attacks, emotionally charged words and pictures, populist generalizations, and logical fallacies.[134]

- **Emotional narratives** (also often mixed with inauthentic or decontextualized images, video or audio) that contain strong personal opinions and that aim to influence the interpretation of certain information – e.g., smearing its source, minimizing the relevance of this information, etc.
- **Images, videos and audio that are fabricated, fraudulently altered or decontextualized,** and which are used to generate confusion, distrust or strong emotions, including through memes that can go viral or false stories. There are many different techniques and modalities within this category, including for instance "**deepfakes**", which are false images or videos created through the use of AI (that can be completely computer-generated or mixed images, video or audio files that already existed). The inauthentic nature of these videos or images can be very hard to detect. Another example of deepfakes is ***synthetic audio***, through which someone's voice is replicated via the use of software that allows for this person's impersonation.[135]
- **Fabricated websites and contaminated datasets** that present false sources or data that have been manipulated, fake websites of governments or companies, or sites that are made to look like news media.[136]

3.4.2. THE TARGETS OF DISINFORMATION WITHIN THE ELECTORAL CYCLE

In the context of elections, the targets of disinformation are usually:

- **Electoral Management Bodies.**
- **Candidates, particularly women candidates.** Research has shown that "false or salacious information about women spreads further, faster and more intensely than disinformation about men",[137] and that disinformation tactics (for instance deepfake videos with sexual content) are often used to shame and deter women candidates and others who aspire to take public leadership positions.[138]
- **Political parties,** which can be the focus of disinformation campaigns aimed at specifically disadvantaging them during electoral periods.
- **Political activists, members of NGOs, CSOs, etc.**
- **Minorities and members of other vulnerable groups,** who are often targeted during elections by disinformation seeking to fuel intolerance and social polarization, including messages focused on their gender identity and expression, sexual orientation, religious, ethnic and racial identities – which can lead to discrimination, hatred and violence.
- **Journalists, media outlets and human rights defenders** that voice critical views, and thus become the target of disinformation efforts aimed at undermining their reputation and discrediting them. Among them, women face gender-specific risks, besides those endured by their male peers.
- **Citizens and voters** in general, through disinformation strategies that manipulate them, generate confusion, and intimidate them, affecting their rights to access to information, freedom of expression, privacy and participation in public affairs.

---

[134] This definition is adapted from the Oxford Internet Institute, https://newsaggregator.oii.ox.ac.uk/.

[135] K. Bontcheva and J. Posetti (eds.), 2020, pp. 22-23.

[136] Ibid., p. 23.

[137] B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: IFES, Working Paper, 'Social Media, Disinformation and Electoral Integrity', August 2019.pdf.

[138] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

By targeting the actors mentioned above, disinformation impacts the electoral process, diminishing public trust in its integrity and the legitimacy of its results, undermining democratic institutions and the social order, generating tensions and violence –all of which could benefit a geopolitical adversary, such as a foreign state or another actor for whom destabilization would be an advantage.

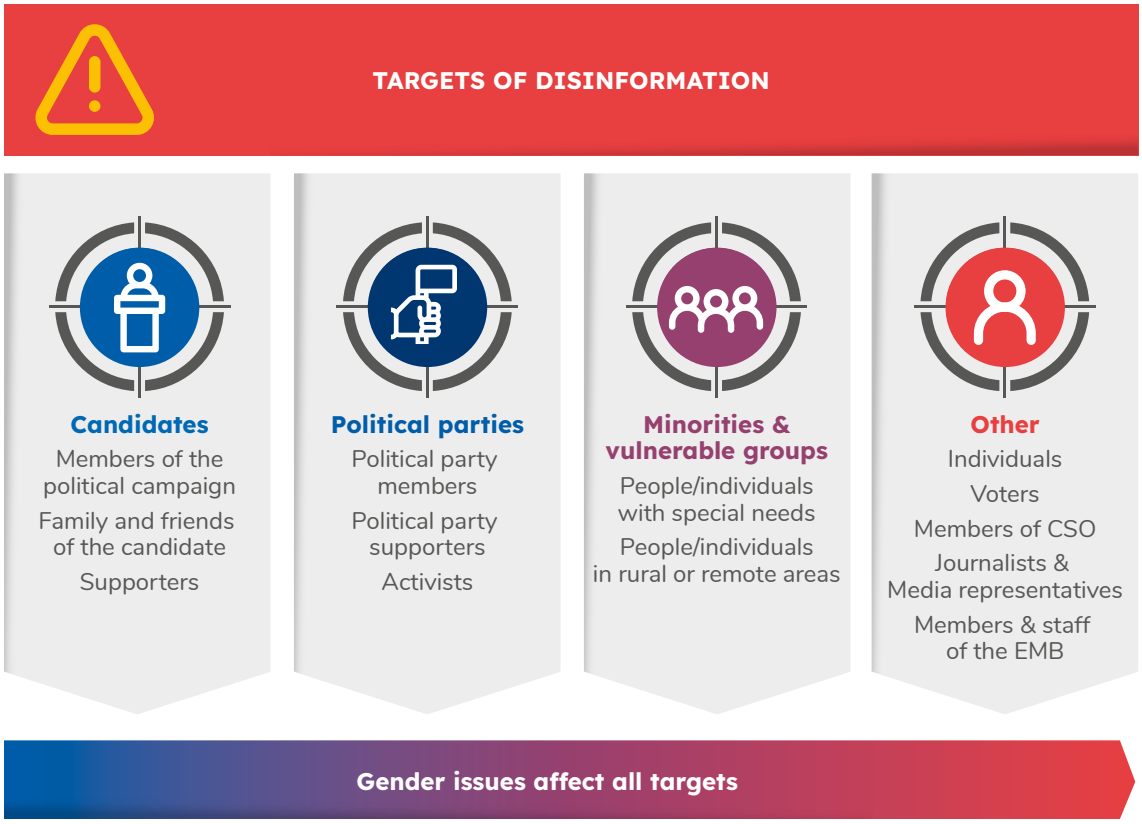


FIGURE 6: TARGETS OF DISINFORMATION

3.5. THE DIGITAL DIVIDE

The digital divide is the gap that exists between individuals who have access to modern information and communication technology and those who lack access.[139] Digital inequality is evident between communities living in urban areas and those living in rural settlements; between socioeconomic groups; between less economically developed countries and more economically developed countries; between the educated and uneducated population. There are numerous types of the digital divide that influence our efforts in accessing the Internet. Some of the vivid gaps in digital inequality include the digital divide between genders.

For elections to be inclusive, women and men must have the same opportunities to vote, choose their representatives and run for public office, without facing unfair obstacles.[140] However, progress in terms of the representation of women in elected and appointed positions falls short of

[139] See: https://www.nngroup.com/articles/digital-divide-the-three-stages/.

[140] UN Women, ***Inclusive electoral processes: A guide for electoral management bodies on promoting gender equality and women's participation***, 2015. Available at: https://www.unwomen.org/en/digital-library/publications/2015/7/inclusive-electoral-processes.

these commitments, despite some improvements in the past years, as noted in the UN Secretary General's 2021 Report on Electoral Assistance.[141]

Despite increases in the number of women at the highest levels of political power, widespread gender inequalities persist: progression in women holding ministerial portfolios has slowed, with just a small increase to from 21.3% in 2020 to 21.9% in 2021; the number of countries with no women in government has increased; and only 25.5% of national parliamentarians are women, compared to 24.9% in 2020.[142]

The pandemic exposed and compounded the obstacles that impede women's full and effective participation and decision-making in public life, further deepening existing inequalities.

The Internet, social media, and AI, particularly in the context of elections, can have an important impact within this priority field of action, including by contributing to sensitize the public about shortcomings in women's political representation and on the importance of efforts to reverse this trend, as well as by facilitating actions to promote women's voting. ICTs can be used to facilitate self-reporting of instances of violence that women experience during elections.[143] Nevertheless, hurdles to access, affordability, (lack of) education and skills and technological literacy, and inherent gender biases and socio-cultural norms, are at the root of gender-based digital exclusion.

Worldwide roughly 327 million fewer women than men have a smartphone and can access mobile Internet. However, women are on average 26% less likely than men to have a smartphone. In South Asia and Africa, for instance, these proportions stand at 70% and 34%, respectively. The gender divide in Internet use is widening. While the global digital gender divide in Internet usage remained almost unchanged between 2013 and 2017, at about 11%, the gender gap in Internet use between developed and developing countries increased, driven by an increase in the gender Internet usage gap by 3 percentage points in least developed countries (LDCs) and 4 percentage points in Africa.[144]

Illiteracy further hinders women's and girls' ability to access online services. About 17% of women worldwide are illiterate, compared to 10% of men (UNESCO, 2017), and illiterate women appear to mainly be using online platform services, such as Skype and YouTube, that are more familiar to them or are easier to access and use. The digital gender divide is also fuelled by digital illiteracy, which often translates into a lack of comfort in using technology and accessing the Internet. Such "technophobia" is often a result of concurrent factors including education, employment status and income level.

Additionally, socio-cultural reasons play an important role in explaining the digital gender divide. In low-income families, it may be privileged for men to own cell phones or have access to the Internet. In very conservative groups, it may be considered inappropriate for girls and women to have access to the Internet or other technologies.

AI technologies have significant gendered implications, including, among others, gender-based

---

[141] Report of the Secretary-General, ***Strengthening the role of the United Nations in enhancing periodic and genuine elections and the promotion of democratization***, 3 August 2021.

[142] The "Women in politics: 2021" map, created by the Inter-Parliamentary Union (IPU) and UN Women, presents global rankings for women in executive, government, and parliamentary positions as of 1 January 2021. The data shows all-time highs for the number of countries with women Heads of State or Heads of Government, as well as for the global share of women ministers. Available at: https://www.unwomen.org/en/digital-library/publications/2021/03/women-in-politics-map-2021.

[143] See: https://www.ifes.org/sites/default/files/vawie_framework.pdf.

[144] Report of the Organisation for Economic Co-operation and Development, ***Bridging the Digital Gender Divide: include, upskill, innovate***, 2018.

exclusion, algorithmic bias and discrimination, the reinforcement of gender stereotypes and misogyny, and the objectification of women.

Yet, social media has also shown to be a useful tool for women politicians. A survey implemented in 2016 by the Women in Parliaments Global Forum,[145] Facebook and the Shorenstein Center on Media, Politics and Public Policy of the Harvard Kennedy School, examined social media use among women parliamentarians in 107 countries. It showed that over 85% of respondents utilized social media, especially for campaigning. The report argues that social media is a political equalizer that is helping women politicians – who are often at a disadvantage regarding their male peers – to break down barriers, given their low entry cost and the equal access they offer for women and men. Relevant in this respect is the survey's finding that women parliamentarians who were members of the opposition or of smaller political parties or groups were the most active on social media platforms. In addition, thanks to social media's flexibility, respondents with children could be as active on social media as those without childbearing responsibilities. However, the survey also showed the new risks brought about by social media, with 50% of respondents reporting that they had received insults or threatening comments questioning women's capacities or roles.[146]

This challenge was also reflected in a 2016 study by the International Parliamentary Union that looked at sexism, harassment, and violence against women in parliament (based on responses by 55 women MPs from 39 countries, 5 regions and 42 parliaments). Respondents noted that psychological violence – particularly through social media – was especially widespread. When asked about the prevalence of different forms of psychological violence they had experienced, 41.8% said that it had entailed the spreading of extremely humiliating or sexually charged images of themselves through social media. Among those surveyed, 44.4% said they had received threats of death, rape, beatings or abduction. Other threats menaced to kidnap or kill their children. These threats were often delivered through email or social media. Many respondents were the targets of online hateful or defamatory comments as well.[147] Women journalists were also victims of online violence and harassment, particularly during electoral periods.[148]

In line with the above-mentioned findings, a 2020 report[149] written by the NGO ShePersisted, described social media as "a double-edged sword for women in politics". On the one hand, it provides them with important avenues for direct communication with the public to deliver an unfiltered narrative in media environments that are still overwhelmingly biased; on the other hand, it exposes them to shocking amounts of sexism, harassment and threats and can amplify gender-based violence (GBV).[150]

---

[145] See: https://www.womenpoliticalleaders.org/wp-content/uploads/2016/10/WIP-Harvard-Facebook-Study_Oct2016.pdf.

[146] See: https://www.womenpoliticalleaders.org/wp-content/uploads/2016/10/WIP-Harvard-Facebook-Study_Oct2016.pdf.

[147] See: http://archive.ipu.org/pdf/publications/issuesbrief-e.pdf.

[148] See: UNESCO, 2021, Journalism is a public good: world trends in freedom of expression and media development: global report 2021/2022; Highlights. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380618?2=null&queryId=0a30ee11-7640-48c0-b1c3-8d7e1e5dc867.

[149] See: #ShePersisted: Women, Politics & Power in the New Media World. The report of the NGO ShePersisted is based on interviews with 88 women in the fields of politics, civil society, journalism, television and technology across 30 countries, and the review of over 100 publications as well as the use of AI to analyse Twitter conversations of men and women candidates in the 2020 Democratic Primary elections in the United States to track gender bias. Available at: https://static1.squarespace.com/static/5dba105f102367021c44b63f/t/5dc431aac6bd4e7913c45f7d/1573138953986/191106+SHEPERSISTED_Final.pdf.

[150] See:https://memo98.sk/uploads/content_galleries/source/memo/fiji/how-women-politicians-on-fiji-treated-on-facebook.pdf.

The study explains that attacks frequently come from coordinated actions of trolls and bots.[151] It identifies widespread gendered disinformation against women political leaders, public figures and journalists, noting that women in politics are the targets of overwhelming volumes of online attacks, fake stories, humiliating or sexually charged images, including photomontages, often aimed at framing them as untrustworthy, unintelligent, emotional/angry/crazy, or hypersexual.[152] These attacks were linked to both their political opponents and to foreign interference. The report warns that online attacks can quickly become offline threats, especially where the rule of law is not strong. Further, it calls attention to the insufficient resources devoted to understanding these challenges and how they affect women's political participation and women journalists.[153]

BOX 4: GENDER-SPECIFIC CHALLENGES FOR JOURNALISTS[154]

Women journalists face gender-based threats, violence, abuse, and harassment, as well as discrimination in their workplace. Attacks include cyberstalking, doxing, rape threats, defamation campaigns, trolling, hacking, and harassment via email, social messaging apps, social media, and digital platforms. Many of these threats and attacks are not reported – given professional, social, or cultural stigma.

The fact that women journalists are more often the targets of online threats and attacks than their male colleagues is documented by several existing studies.[155] The nature of the digital harassment and abuse they face is also different, as it generally includes sexual references, sexist language, allusions to their physical features, personal relationships, or cultural background.[156]

In 2020, UNESCO and the International Center for Journalists (ICFJ) launched a global study to assess the scale of online violence against women journalists and analyse good practices to address this problem.[157] The results of the online survey showed that 73% of the journalists identifying as women who responded to the survey had experienced online violence while performing their job, 25% had been the targets of threats of physical violence, 18% had received threats of sexual violence and 20% had been targeted by offline attacks connected to the online violence they had experienced. In terms of the impacts of these threats and violence, 30% stated that they had self-censored and 26% responded that their mental health was affected. In 41% of the cases of online attacks, these were seemingly connected to orchestrated disinformation campaigns.[158]

---

[151] See: https://www.she-persisted.org.

[152] Ibid.

[153] See: https://www.she-persisted.org.

[154] See also: J. Posetti, N. Shabbir, D. Maynard, K. Bontcheva and N. Aboulez, ***The Chilling: Global trends in online violence against women journalists***, UNESCO, 2021. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000377223.

[155] See: ibid. and UNESCO, 2021, ***Practical guide for women journalists on how to respond to online harassment***, Paris. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000379908.

[156] See: https://unesdoc.unesco.org/ark:/48223/pf0000371487?fbclid=IwAR2FAIrs5INt5ibUb_gHxfNYaFEzTYlTbMeEo-b8ZjXfjbDFAII1ad8vHAgk.

[157] See: https://en.unesco.org/news/closed-global-survey-online-violence-against-women-journalists-and-effective-measures-combat; https://unesdoc.unesco.org/ark:/48223/pf0000375136.

[158] See: https://en.unesco.org/news/unescos-global-survey-online-violence-against-women-journalists.

ACTIVITY III

The following activity has the objective of determining the participant's level of knowledge of the new information paradigm; the concepts of disinformation, misinformation and mal-information; the drivers behind misleading content; who instigates, produce and spreads disinformation and how; which are the different tactics and technics; who are the targets of disinformation during and within the electoral cycle; the digital divide and the gender divide.

Suggested guiding questions for a discussion:

I. Which are the main characteristics of the new information paradigm?
II. Please provide some examples of disinformation and misinformation. Can you explain the risks of simply 'liking', 'sharing', and commenting on posts where you have not ascertained whether they are likely to be true or not?
III. What are the drivers behind the production and distribution of misleading content? Can you give an example?
IV. Why can the tactics and techniques to spread disinformation affect freedom of expression and the electoral processes? Please provide some examples.
V. Who are the targets of disinformation during the electoral cycle?
VI. Please provide some examples of how women can be affected by the digital divide in the context of elections (women candidates, women journalists, electoral practitioners, etc.).

4. IMPACT OF SOCIAL MEDIA AND AI IN THE ELECTORAL CYCLE

OBJECTIVES OF THIS SECTION

- Understand the impact of social media platforms on citizens during electoral processes.
- Assess how algorithms can amplify misleading content.
- Explore the concepts of voter suppression, Internet shutdown and disruption of Net neutrality via zero-rating.
- Examine the critical role of cybersecurity, digital campaigning and how technology contributes to increasing violence against women and journalists during elections.

Social media platforms allow political contestants to better reach out to their voters and to engage them more directly in their campaigns, without involving the traditional intermediary role provided by traditional media. They also enhance the opportunities for citizens to retrieve information that is important for their voting decisions, which is particularly impactful where freedom of expression and access to information is restricted – where social media networks often provide the only means for opposition candidates to communicate their views to voters.

Social media networks help individuals with alternative ideas to connect and be heard in countries where significant media groups are owned by a few influential people with their political agendas. Moreover, social media networks serve as very efficient and relatively cheap tools for voter mobilization. Nevertheless, recent electoral processes showed that social media networks can also be used for harmful purposes that, in many cases aided by AI technologies where algorithms amplify engagement of emotive and sometimes false content, undermine the integrity of elections all along the electoral cycle.

More specifically, social media networks have been used for:

- **Voter suppression** (misrepresentation of «factual» voter information – e.g., regarding methods, place, location, time, qualifications, and identification).
- **Voter fraud** (vote buying/selling).
- **Incitement to violence, spreading of hate speech** (to heighten deep-seated sources of tension, discord, and hatred - including calls for political and electoral exclusion - in ways that undermine public trust in democratic institutions and increase the risk of electoral violence and political instability).
- **Bullying, harassment and arbitrary surveillance** of activists, candidates, journalists, or other public figures on public areas of social media, as well as via private social messaging targeting the victims.
- **Cyberespionage** (a form of cyberattack that steals classified or sensitive data to undermine a candidate or party).
- **Doxing** of candidates and activists for the purposes of harm, harassment, online shaming, manipulation and shaping the opinions of voters (doxing means publishing personal information about people without their consent, which can include addresses, phone numbers, credit card details, medical information, private e-mails, etc.).
- **Data mining for micro-targeting** used in electoral campaigns (political micro-targeting is even more tailored to individual voters than normal advertising used for example on television, and relies on a broad set of collected data about an individual, based on traces the person leaves through online interactions), with the effect that only selective political messages reach potential voters, thereby providing a misleading rendition of the comprehensive platform being advanced by a political contender.
- **Spreading disinformation** (false or misleading information that is created or disseminated

with the intent to cause harm – to persons, groups, institutions or processes – or to benefit the perpetrator).[159]

- **Exerting foreign interference in elections** (attempts by governments, covertly or overtly, to influence elections in another country).
- **Trolling** (aimed at generating online discord by upsetting people or starting quarrels, through content that is inflammatory or off-topic that is posted in an online community. Trolls can for example be hired by political contestants – parties or candidates – to discredit opponents).
- **Identity theft** (the deliberate use of someone else's identity to discredit opponents, for example by stealing personal emails and other data).
- **Digital attacks against women, journalists and other political, institutional and media actors.** This includes attempts to limit legitimate political speech through the shutdown of the Internet and other communications channels, filtering or blocking content pertaining to elections, applications and websites, as well as illegitimate surveillance, tracking, hacking, fake domain attacks, denial-of-service (DoS) attacks, data mining, doxing, confiscation of digital hardware, among other modalities.

4.1. ELECTIONS, DISINFORMATION AND CONFLICT PREVENTION

Electoral processes offer a safe, predictable, rule-bound method for arbitrating political and social conflicts through the selection of representatives or the definitive resolution of questions before the community (as in referendums). When elections are credible, they strengthen the capacity of the State to ensure security through legitimate authority under the rule of law, and to improve levels of human development through effective governance. Credible, periodic, and transparent elections are the basis of legitimate governments that enjoy popular support for programs and policies.[160]

On the other hand, precisely because electoral processes are contests through which political power is retained or pursued and social differences are highlighted by candidates and parties in campaigns for popular support, they can often generate vulnerabilities for the escalation of conflict into violence.[161] Tensions may arise in the run-up to election processes as some candidates mobilize along extremist lines to win support, as rival factions vie for votes and to secure turf, and as parties or factions seek to weaken or even eliminate opponents in efforts to seek or retain political power.[162]

During the election event, as well, violence (including violence against women), might spike in the days before or during voting as the drama of the contest unfolds. After the vote, there is the continuing potential for post-election violence when allegations of fraud and corruption emerge, or when those dissatisfied with the outcomes of elections take to the streets or, in the gravest instances, the battlefield, to challenge results. Thus, electoral processes can contribute to peace, or they can be catalysts of conflict. In this context, disinformation, misinformation and mal-information can contribute decisively to the perpetration of violence undermining the legitimacy of the electoral process and of the bodies in charge of it.

Credible elections are based also on free and equal access to good quality information and hence extremely vulnerable. Preventing violence related to electoral processes nowadays needs to

---

[159] L. Reppell and E. Shein, ***Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions***, International Foundation for Electoral Systems, 2019. Available at: https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf.

[160] ***Elections and conflict prevention: a guide to analysis, planning and programming***, UNDP, Democratic Governance Group, Bureau of Development Policy, 2009. Available at: https://aceproject.org/ero-en/misc/elections-and-conflict-prevention-a-guide-to/view.

[161] Ibid.

[162] Ibid.

consider the new information paradigm, which acts as a generator and amplifier of conflict that could delegitimize democratic processes.

4.2. ELECTIONS AND CYBERSECURITY

Effective cybersecurity plays a critical role in the EMBs operational planning. Elections rely on varying combinations of manual and technology-based procedures. As neither truly "unhackable" technology nor entirely tamper-proof manual processes exist, an essential task in election administration involves the management and mitigation of manipulation risks through a range of integrity, audit and control measures. With the expanded use and dependency on ICTs, the dangers of interference in and manipulation of democratic electoral processes also grow.[163] This requires responding to constantly evolving challenges and cyber threats that all stakeholders need to be aware of and equipped to address and prevent.

In this context, cybersecurity relates to protecting Internet-connected systems, networks, software, and data from unauthorized access of exploitation, including the security of offline election technologies, and protecting the integrity of the electoral process from hacking, disinformation and influence operations.[164] The field of cybersecurity is complex and cross-cutting, encompassing strategic, technical, legal and security issues, and requires collaboration across sectors.[165] As cyber-fuelled attacks can critically undermine the legitimacy of elections and the mechanisms to protect them, safeguarding the technology involved in the electoral process is key.[166] Towards this aim, the Network and Information Systems (NIS) Cooperation Group drafted a Compendium on Cyber Security of Election Technology[167] that identifies the following cyber threats all along the electoral cycle:[168]

During the pre-electoral period:

- **Threats related to party/candidate registration:** tampering with registrations; denial-of-service attacks (DoS) or overload of party or campaign registration, resulting in them missing the deadline; fabricated sponsor signatures.
- **Threats related to electoral rolls:** identity fraud during voter registration; deletion of or tampering with voter data; DoS or overload of the voter registration system, suppressing voters.
- **Threats related to the campaign's information technology:** hacking of candidates' laptops, email accounts or social media accounts; hacking of campaign websites (e.g., through defacement, DoS); website misconfiguration; leak of confidential information.
- **Threats related to public communication, the media:** hacking of internal systems used by the media; tampering, DoS, or overload of media communication links; defacement, DoS, or overload of websites or other systems used for the publication of the results.

---

[163] Freedom House, ***Report on the Crisis of Social Media***, pp. 7 and 8 on The Global Phenomenon of Digital Elections Interference and the Key Tactics of Digital Election Interference, 2019. Available at: https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf.

[164] S. Van der Staak and P. Wolf, ***Cybersecurity in Elections, Models of Interagency Collaboration***, p. 9, IDEA, Stockholm, 2019. Available at: https://www.idea.int/publications/catalogue/cybersecurity-in-elections.

[165] See: https://www.itu.int/en/ITU-D/Cybersecurity/Pages/default.aspx.

[166] NIS Cooperation Group, ***Compendium on Cyber Security of Election Technology***, 2018. Available at: https://www.ria.ee/sites/default/files/content-editors/kuberturve/cyber_security_of_election_technology.pdf.

[167] See: https://www.ria.ee/sites/default/files/content-editors/kuberturve/cyber_security_of_election_technology.pdf -The NIS Cooperation Group is constituted by representatives of EU Member States, the European Commission and the European Union Agency for Network and Information Security, and was established by the Directive (EU) 2016/1148 'concerning measures for a high common level of security of network and information systems across the Union' (NIS Directive). It facilitates strategic collaboration between EU Member States regarding network and information systems' security.

[168] Ibid., p. 17.

During the electoral period:

- **Threats related to voting election technology:** tampering or DoS of voting and/or vote confidentiality during or after the elections; software bug that alters election results; tampering with logs or journals; breach of voter privacy; tampering, DoS, or overload of the systems used for counting or aggregating results; tampering or DoS of communication links used to transfer (interim) results; tampering with the supply chain involved in the movement or transfer of data.

During all phases:

- **Threats related to the government's information technology:** hacking or misconfiguration of government servers, communication networks, or endpoints; hacking of government websites or social media accounts, spreading misinformation on the election process, registered parties/candidates, or results; DoS or overload of government websites.

In recent elections around the world, there have been numerous examples of the use of information that is false, misleading, contradictory, and exaggerated with the purposes of amplifying voter confusion, undermining fact-based political debate, attacking candidates' reputations, deepening social tensions, mobilizing supporters, restricting activism, incrementing violence, and marginalizing women and minorities, among others.[169]

4.3. DIGITAL CAMPAIGNING

Many political actors worldwide are increasingly using digital and social media platforms to campaign, especially by buying advertising services from companies like Meta, Google, Twitter or TikTok.

Digital campaigning allows candidates to reach new voters, which is positive for electoral participation. It can also make it easier and cheaper for campaigners to communicate with citizens, explain their policies and political views. On the other hand, new techniques for reaching voters– including micro-targeting, data mining, data harvesting, and the creation of psychometric profiles – could reduce confidence in the integrity of elections.

Moreover, current regulatory frameworks tend to lag in this field. In many cases, they do not cover online political advertisements, and transparency regarding them is not guaranteed through standard reporting requirements. Experience from recent elections shows that, in this context, it is easier for foreign individuals or States to try to influence voters online without the need to have any physical presence in a country. It is also possible for campaigners to try to get around limits set on spending through unreported digital advertising. Another critical problem is the misuse of private information by political campaigns, parties, social media companies, and other commercial organizations.

There is a need to enhance the transparency of online political advertising, including its source of financing, targeting methodology, and levels of funding, as well as the accountability of technology companies to national legislatures and other regulatory organs. A collaborative approach is necessary at the international level concerning regulation. Now, the tendency is for national-level regulation to not adequately guide technology companies and advertisers on good practices in the field of digital political advertising, but that is likely to change.

4.3.1. MICRO-TARGETING AND THE USE OF DATA

As a new form of political advertising, micro-targeting typically involves monitoring people's online behaviour, and using the collected data, sometimes enriched with other data, to display

[169] B. Martin-Rozumilowicz and R. Kužel, ***Social Media, Disinformation and Electoral Integrity***, IFES working paper, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.

individually targeted political advertisements. However, micro-targeting poses serious risks, as demonstrated by the Cambridge Analytica scandal, where a voter-profiling company had harvested private information from the Facebook profiles of more than 50 million users without their permission.[170]

Unlike political advertising on television, micro-targeting not only affects the democratic process, but it also affects people's privacy and data protection rights. Indeed, micro-targeting affects myriad other rights and duties, including a political party's and online platform's right to impart information, a voter's right to receive information, and the government's duty to ensure free and fair elections.[171] Its impact is also linked to the legal framework of each country and region, i.e., the rules for the protection of privacy and personal data.

Political micro-targeting (PMT) is an election campaigning approach that uses different communication methods (e.g., email, phone, canvassing, social media, etc.) to gain the voters' trust.[172] It aims to build a relationship with the electorate to generate a favourable voting outcome. Micro-targeting can be defined as a "marketing strategy that uses people's data – about what they like, who they are connected to, what their demographics are, what they have purchased, and more – to segment them into small groups for content targeting."[173] Profiling for political micro-targeting purposes entails essentially the same kind of operation, whereby voters' data is used to split them into small, niche groups, to aim them with tailored messages according to their characteristics – such as their psychometric profiles.[174]

Unlike traditional advertising, micro-targeting shows specifically curated content to targeted groups of voters grouped in the same cluster according to their concerns, opinions, and beliefs. Candidates and parties can make different promises or even contradictory claims to diverse groups of voters with little oversight or accountability. PMT is not only a tool to mobilize political participation but can also be used to discourage it, as well as to encourage individuals to abstain from political donations.[175]

Micro-targeting relies on AI, based on algorithmic processing of the digital traces a person leaves when interacting online.

BOX 5: HOW DOES AI OPERATIONALIZE MICRO-TARGETING?

AI systems consist of three steps: perception/sensing, reasoning/processing and actuation.[176] The sensing step is enabled by input devices such as cameras, microphones, keyboards or websites. In the context of elections, large data volumes is taken in and used to create personalized profiles of members of the electorate based on their preferences and antipathies. The key function of any AI system happens during the processing

[170] C. Cadwalladr and E. Graham-Harrison, ***Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach***, The Guardian, 2019. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

[171] T. Dobber, R. Ó Fathaigh and F. J. Zuiderveen Borgesius, The regulation of online political micro-targeting in Europe, ***Internet Policy Review***, 2019.

[172] B. Bodó, N. Helberger and C. H. de Vreese, 2017, Political micro-targeting: a Manchurian candidate or just a dark horse?, ***Internet Policy Review***, 6(4).

[173] See: https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/.

[174] Report by Panoptykon "Who (really) targets you?" - Facebook in Polish election campaigns. Available at: https://panoptykon.org/political-ads-report.

[175] Ibid.

[176] EC (Venice Commission), 2019, ***Joint Report of the Venice Commission and of the Directorate of Information Society and Action against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections***, CDL-AD (2019)016.

stage, where the data received through sensing is processed, and recommendations are made according to predefined goals. In the case of micro-targeting, the profiles are checked against the previously set objectives. Corresponding content is created. The AI tool decides – based on the sensed profiles – what content to distribute and or what behaviour would be appropriate to achieve the pre-set goals: for example, to shape one's opinion about a particular election candidate or party.

After the decision-making step, the machine steps forward to actuation, where it will act according to the decision that was made. In the context of political micro-targeting, this would be the phase in which the content that seems most appropriate given the pre-set goals is distributed, in the form of tailor-made micro-messages or online engagement more generally (a like, a post, a share, a reaction, etc.). This action will later be perceived by the sensors of the machine as an alternation, and all information will be processed again in order to decide upon the next online activity, i.e., how and with what content to micro-target the audience next.

The accessibility of big data might be the most significant driver of the recent rise of AI's importance, and it underpins its role in election campaigning.[177] Big data is defined as the accessibility, at a high velocity, of high volumes and a high variety of data sets, which require cost-effective and innovative processing methods that permit enhanced decision-making and process automation.[178] The processing of big data covers large quantities of unfiltered and unstructured data and the logging of all kinds of online events and interactions such as posts, updates, reactions, direct messages, re-posts, etc.[179]

Data mining and data harvesting

When examining the issue of political micro-targeting, the concepts of data harvesting and data mining should also be considered. These are two complementary data analytics strategies, yet there are some important differences between them.[180]

- **Data harvesting** refers to a more recent trend in data analytics strategies and describes the process of large data extraction from online sources through the use of a malicious bot.[181] It is a cheap method for obtaining a large amount of data without the permission of its holders.
- **Data mining** refers to the process of analysing large data sets to detect patterns, relationships, links, or trends.[182]

With these sophisticated data analytics strategies, campaigning parties can detect and analyse voters' behaviours and responses on certain political issues and events, and consequently make

[177] J. Berryhill, K.K. Heang, R. Clogher and K. McBride, 2019, ***Hello, World: Artificial intelligence and its use in the public sector***, OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, https://doi.org/10.1787/726fd39d-en; C. J. Bennett and D. Lyon, 2019, Data-driven elections, ***Internet Policy Review***, Vol. 8, No. 4.

[178] Gartner, 2020, Big Data. Available at: https://www.gartner.com/en/information-technology/glossary/big-data. See also: R. Krimmer, G. Hammerschmidt, T. Husted, M. Kleinaltenkamp, S.J. Mikhaylov, C. Raffer and C. Schmidt, 2021, ***Good-Practice-Beispiele der Digitalisierung öffentlicher Verwaltung im Ausland***. (Forthcoming).

[179] E. K. Jaisal, 2018, ***Data-mining and Analytics: Rising Concerns over Privacy and People's Security***. Available at SSRN 3472729.

[180] Import.io, 2019, ***Data Mining Process: The Difference Between Data Mining & Data Harvesting***. https://www.import.io/post/the-difference-between-data-mining-data-harvesting/.

[181] Caspio, 2014, ***What You Need to Know about Data Harvesting and How to Prevent it***. Available at: https://blog.caspio.com/what-you-need-to-know-about-data-harvesting-and-how-to-prevent-it/; Import.io, 2019.

[182] Import.io, 2019.

better decisions on how to react further on, as well as generate tailor-made micro-messages.[183] Data harvesting and data mining are very effective in creating complete psychometric profiles of voters based on the content they post, share, or react to on social media platforms, in private messages, as well as their search histories.[184]

Psychometrics is a scientific discipline that examines people's psychological traits to categorize the distribution of five big personality traits within society: agreeableness, extraversion, openness to experience, conscientiousness, and emotional stability/neuroticism.

According to these traits, personalized profiles are created, containing information based on gender, sexual orientation, race, religion, political beliefs, relationship status, potential addictions, peer groups, and other important, often very intimate information.[185] Profiling also relies on meta-data, which relates to context – such as where a potential voter was when they posted on social media, for how long, and why. Meta-data can be attained, for example, through the location services of any smartphone, which is in constant contact with Wi-Fi routers or cell towers and thus allows companies to gather this valuable information.[186]

Individuals can be targeted with information that is more relevant to them, but also with manipulative content according to what their personalities are likely to be more receptive to.[187] In consequence, the electorate might be influenced into making decisions that deviate from their original political ideas, without even realizing it.[188] The use of psychometric profiles has a high chance of yielding positive results, as it triggers individuals' emotions, and it has made campaigning more cost-effective.[189] Political micro-targeting (PMT) can take various forms and use different channels. For example, it can be paid advertising that appears in a person's social media feed. The data that serves such targeting can also be purchased or accumulated to inform direct communications by political actors and/or supporters, through individual messages delivered through public or private means. Beyond deliberate targeting, algorithmic systems can rank content for users based on a range of variables, thereby reinforcing beliefs and interpretations.

BOX 6: KEY CHARACTERISTICS OF POLITICAL MICRO-TARGETING

Political micro-targeting (PMT) is a highly effective tool that can both significantly contribute to, and that relies on, voter modelling and that allows candidates to approach specific groups of the electorate individually. At the same time, PMT can heighten division and polarization in society and fragment political discourse.

Filter bubbles and echo chambers emerge as issues of concern, given that the electorate only receives online content that is presented according to their presumed or prior proven interests/online behaviour.[190] It tends to provide content to certain individuals on topics that are important to them, rather than confronting them with diverse opinions.

---

[183] E. K. Jaisal, 2018

[184] T. Blesik, M. Murawski, M. Vurucu and M. Bick, 2018, Applying big data analytics to psychometric micro-targeting, in ***Machine Learning for Big Data Analysis (Vol. 1)***.

[185] Susser, Roessler and Nissenbaum, 2018.

[186] J. Chester and K. C. Montgomery, 2019, The digital commercialisation of US politics—2020 and beyond, ***Internet Policy Review***, 8(4), 1-23.

[187] T. Blesik et al., 2018.

[188] Susser et al., 2018.

[189] Ibid.; E. K. Jaisal, 2018.

[190] S. Barocas, 2012, ***The price of precision: Voter microtargeting and its potential harms to the democratic process***. Paper presented at the Proceedings of the first edition workshop on Politics, elections and data.

As operationalized through digital platforms and social media, PMT might act beyond the legal period established for political campaigns.[191]

PMT can benefit candidates with less financial capacity, by allowing them to efficiently focus and reach voters. Nevertheless, political micro-targeting based on psychometric profiling requires significant resources and is not easily conducted.

Companies that manage this technology might have conflicts of interests, a fact that could undermine the principle of impartiality and equality of opportunity for the contenders in an election.[192]

Two prominent companies that hold vast amounts of datasets regarding individuals are Meta and Google.[193] The latter can recreate a digital clone of each user and might be able to predict their decisions better than themselves. By using highly sophisticated and innovative tools, they can enhance their advertising capabilities through personalization techniques that create related advertisement features.[194] As the business of using AI techniques to harvest personal data for micro-targeting purposes has grown, social media companies and Internet intermediaries often team up with marketing agencies to improve the understanding of the content people are most susceptible to, to increase the appropriate advertisement. This commercial strategy has had a significant impact on electoral campaigns.

BOX 7: THE USE OF CHATBOTS AND FAKE ACCOUNTS FOR POLITICAL PURPOSES

A common strategy to market political interests and advertise candidates is to deploy *chatbots*, that is, computer-based agents that interact with users via automated accounts that appear to be human. In the political sphere, the so-called *political bots* are often used for advertising political messages and propagandistic content. In combination with the method of profiling people according to their personality traits, data analytic companies use political bots which selectively target individual voters with emotional messages to manipulate their decision-making.[195]

Another campaign strategy nowadays includes creating fake social media accounts to produce the illusion that candidates are more popular than they are to win more authentic followers.

[191] J. Bennett and D. Lyon, 2019.

[192] T. Blesik et al., 2018; I. Nenadić, 2019, Unpacking the "European approach" to tackling challenges of disinformation and political manipulation, ***Internet Policy Review***, 8(4), p. 3.

[193] E. K. Jaisal, 2018.

[194] J. Chester and K. C. Montgomery, 2019.

[195] B. Anderson, 2017, ***The Rise of the Weaponized AI Propaganda Machine***. Available at: https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b.

BOX 8: MESSAGING SERVICES AND CLOSED GROUPS

The use of SMS or private messaging services such as WhatsApp, Messenger, Signal and Telegram has critical implications in elections. These services are very popular and they allow one-to-one communications as well as closed groups interactions mostly for free.

Messaging services and closed groups have also become vectors of disinformation and inflammatory speech throughout the electoral cycle. Political advertising is distributed on closed networks and allows political campaigns to avoid scrutiny from regulators and reporters.

The advantages of using this kind of platform for campaigning are clear: not only does it allow strategists to tailor messages to various interest groups, but it also protects their anonymity.

The end-to-end encrypted nature of these services, coupled with the privacy they provide, is a double-edged sword for democracies. While private messaging services facilitate the sharing and distribution of legitimate content (including civic and voter education, candidates' proposals, etc.), they also make it easier to influence the electorate through misinformation and disinformation. The potential to manipulate and alter information, use technology to deceive uninformed users, and circumvent control mechanisms can eventually undermine the integrity and credibility of the electoral processes.

4.3.2. INFLAMMATORY LANGUAGE IN THE ONLINE ENVIRONMENT

The visibility and reach of hateful and violent content – particularly which is directed towards vulnerable groups – have increased considerably in the last few years through its circulation on social media networks. The proliferation of these forms of expression, which has had a chilling effect and undermined the affected communities' online engagement, represents a significant challenge for policymakers to tackle.

Hate speech online has been described as the intersection of multiple tensions, as it is the "expression of conflicts between different groups within and across societies; it is a vivid example of how technologies with transformative potential such as the Internet bring with them both opportunities and challenges; it implies complex balancing between fundamental rights and principles, including freedom of expression and the defence of human dignity".[196]

Every electoral campaign has one goal: to convince the electorate to vote for a particular candidate or party. Political campaigns provide fertile ground for exaggerated promises, over-blown statements of facts, lies, and misrepresentations. With the rise in the use of social media to spread campaign messages, hate speech has become one of the most pervasive problems.

In 2019, 26 UN mandate holders deplored the rise of hate speech through a joint statement in which they stated that it "has become mainstream in political systems worldwide."

They expressed concern, particularly for vulnerable groups as leaders, government officials, politicians, and other prominent public figures sometimes spread discriminatory and hateful content for their political gain.

The UN calls "on public officials and politicians, as well as the media, to assume their collective

[196] UNESCO, 2015, ***Countering Online Hate Speech***, p. 7. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000233231.

responsibility to promote societies that are tolerant and inclusive". Furthermore, it highlights those efforts to curb hate speech should not restrict freedom of expression but just the opposite, as "freedom of expression serves as a vital tool to counter hate speech".[197]

Disinformation campaigns that use hate speech as a tactic rely and build on underlying social dynamics and existing divisive messages and affinity groups.[198] To prevent this trend, it is necessary to identify the different components involved in the generation of hate speech and disinformation online:

- **Actors:** in general, individuals that produce hate speech are either motivated by ideological factors, their personal beliefs, or views of the world. Nevertheless, hate speech is increasingly generated by coordinated actors (either domestic or foreign) who aim to undermine political participation, cause confusion, distrust, and polarization; to harm social cohesion and democracy, support a person or group or make money.
- **Messages:** may vilify, humiliate, or promote intolerance and violence against groups of persons by explicit reference to their race, national or ethnic origin, religion, gender, sexual orientation, age, disability, or other shared identity. Also, inauthentic content amplifies narratives that are already in circulation, relying on the cognitive biases of the people who engage with them, as well as on the manipulation of content and images through AI.
- **Mode of dissemination:** hate speech can turn viral on traditional media and social media platforms, even in the absence of a disinformation campaign. Inflammatory or inauthentic content is spread through human, bot and hybrid paid engagement, gaining an appearance of credibility – which makes it more likely to be reshared as well as to be circulated through traditional media and through word of mouth. Personal data powers algorithms and AI to micro-target messages, making them more persuasive to specific audiences.
- **Interpreters:** the threats that hate speech poses to political/electoral processes and institutions derive from how citizens receive and interpret hate speech and disinformation. Hate speech messages appear to be more popular given their manufactured amplification. This may embolden citizens that are normally passive. It can also impact the level of hostility against certain groups or on the electorate's perception of popular opinion. Citizens may perceive that the integrity of political and electoral processes is threatened.
- **Risks:** The result of these strategies and the circulation of problematic content may imply the risk of delegitimizing the electoral process and thus weakening democratic institutions. Hateful, false content might contribute to diminished trust in democratic institutions and processes, exclusion of the targeted groups from democratic engagement and violence.

It should be clarified, however, that although hate speech and disinformation often intersect, they are two different concepts. In turn, hatred does not always involve disinformation either, but can amplify fears, misogyny, xenophobia, and other prejudices through expressions of opinion and incitement that do not necessarily involve falsehoods.[199] In particular, hate speech (including sexist hate speech) against women electoral stakeholders – including candidates, political party leaders, activists, journalists, voters, and electoral officers – can have a deep negative impact on women's political participation and women's rights.

---

[197] Joint open letter on concerns about the global increase in hate speech signed by 26 mandates. Available at: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25036&LangID=E.

[198] L. Reppel and E. Shein, 2019. Available at: https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf.

[199] K. Bontcheva and J. Posetti (eds.), 2020, ***Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression. Broadband Commission research report in 'Freedom of Expression and Addressing Disinformation on the Internet'***, p. 31. Available at: https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf.

4.4. INTERNET SHUTDOWNS AND ARBITRARY THROTTLING

An Internet shutdown is an intentional and significant disruption of the Internet or electronic communication ordered or sanctioned by the authorities and sometimes targeted at a predetermined group of people. It can restrict access to specific social media platforms, mobile applications, or the Internet as a whole.[200] In turn, arbitrary throttling refers to the means of deliberately reducing the speed of the Internet by a service provider.[201]

The use of Internet shutdowns during elections contradicts the vision set out in UN Resolution A/HRC/RES/32/13, which was adopted to promote, protect, and advance the enjoyment of human rights on the Internet.[202] In their 2018 Joint Declaration, the Special Mandates on Freedom of Expression recalled that while "generally unacceptable under international law", measures such as Internet shutdowns and other similar ones are particularly problematic "in the context of political debate and elections".[203] The 2011 report of the then UN Special Rapporteur Frank La Rue underlines that "the Internet, as a medium by which the right to freedom of expression can be exercised, can only serve its purpose if States assume their commitment to develop effective policies to attain universal access to the Internet."[204]

In 2019, the number of Internet shutdowns worldwide amounted to 182 documented cases in over 34 countries, according to the NGO Access Now. Many Internet shutdowns and instances of arbitrary throttling take place during elections, endangering voter mobilization and access to information.

Shutdowns can jeopardize democratic integrity, particularly if the agents who provoke these situations are part of the political party or government in power. Governments often justify Internet shutdowns or throttling during electoral periods by referring to the fast-spreading of disinformation, conspiracy theories, and hate speech, or by citing the need to prevent fraud or election-related violence fuelled by organized mass protests. In some contexts, the justification given for resorting to an Internet shutdown is the goal of preventing the dissemination of false or early of election results.[205]

Internet shutdowns restrict freedom of expression, impede the timely communication of election results, and limit other activities by election observers. They can affect the ability not only of electoral monitors, but also of citizens, to report instances of fraud or other incidents. Shutdowns impact on the work of media actors, curbing their capacity to disseminate information about the electoral process. They can also affect the role of institutions such as the judiciary, CSOs and political support groups.[206] Given the media coverage restrictions resulting from these shutdowns, it is not rare for elections in which these occur to give rise to human rights abuses by security forces.

---

[200] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

[201] B. Taye, 2019, ***TARGETED, CUT OFF, AND LEFT IN THE DARK. The #KeepItOn report on internet shutdowns in 2019***. Available at: https://www.accessnow.org/cms/assets/uploads/2020/02/KeepItOn-2019-report-1.pdf.

[202] HRC, 2016.

[203] See: https://www.osce.org/files/f/documents/1/e/379351.pdf, p. 2.

[204] Report of Special Rapporteur on the promotion and protection of the right to freedom of expression and opinion, 16 May 2011 (A/17/27).

[205] UN, 2019.

[206] AccessNow, 2016, https://www.accessnow.org.

4.5. ARBITRARY BLOCKING AND FILTERING OF ONLINE CONTENT

Arbitrary blocking or filtering of online content are two forms of Internet censorship that can also take place all along the electoral cycle, altering democratic processes. Some governments are utilizing their power over the Internet's physical infrastructure and on Over-the-Top online services (OTTs) to block or filter content in order to steer the election outcome in a way that favours their own interests, or as a counter-measure to protests that might occur during or after elections.

Arbitrary blocking or filtering of online content can have a variety of different faces. One method is to utilize denial-of-service attacks, which make certain websites inaccessible.[207] Another method is Domain Name System tampering, which refers to blocking domain names. Another technique is HTTP(S) blocking, which change HTTP(S) search entries into entries that direct the user to a blocked page.

BOX 9: DETECTING ILLEGITIMATE BLOCKING OF CONTENT AND OTHER ONLINE DISRUPTIONS

Several global initiatives have emerged in opposition to the practices analysed in this section. One of them is the Internet Shutdown Observatory created by the NGO NetBlocks, which detects online disruptions, blocked content, censorship, or cyber-attacks on critical infrastructure. NetBlocks is a London-based organization that exists since 2016. Its operations cover Asia, Africa and Latin America.[208]

Detecting unwarranted blocking of content or censorship is not a linear practice and is often hard to implement. It requires a comprehensive analysis to verify whether a disruption was intended to restrict access to Internet content, or whether it was a by-product of other legitimate measures taken by the government. NetBlocks has a combined approach to monitor incidents.[209] Firstly, it uses analytical software to monitor whether and how millions of mobile devices can access certain online content. Secondly, NetBlocks closely cooperates with local OTTs and engineers that help confirm whether an incident that occurred is related to a deliberate Internet content restriction or might have been caused by other factors. This method serves to rule out the need to carry out further inquiries regarding certain Internet disruptions, thus allowing NetBlocks to focus on the ones that appear to be government-created.

The collected data obtained from this combined method is used to train a machine learning algorithm to produce increasingly automated detection and analytics tools to be applied to future events.

[207] R. Deibert and R. Rohozinski, 2010, ***Access controlled: The shaping of power, rights, and rule in cyberspace***.

[208] NetBlocks, 2020b.

[209] Hsu, 2020.

4.6. DISRUPTION OF NET NEUTRALITY VIA ZERO-RATING

The disruption of Net Neutrality (NN) refers to the principle that all OTTs "should treat all data equally and not prioritize data or services for any reason – including commercial and political ones".[210] NN is important for the rights to freedom of expression and access to information because it protects a person's "choice and right to access Internet content, applications, services and hardware". Because digital content travels through multiple layers of the "tech stack", it is possible for blockages to be implemented at various levels – from operating systems, OTTs, network security providers, cloud storage providers and app stores, and not only through content-layer applications such as Meta, Twitter, Google, RenRen.com or Telegram among other.

Internet providers have the technical capacity to discriminate against certain content or services by either restricting them or simply regulating the speed of Internet traffic.[211] One way of doing so would be to favour content delivery by certain providers and to slow content delivery by others. Net Neutrality can be disrupted through a strategy called zero-rating, which refers to the:

> "...discriminatory technique where telecom operators allow customers access to select online content or services at no additional cost through a prior arrangement with content providers. The selected sites are rated at zero cost to the customers, violating the essence of net neutrality, which requires non-discrimination between different content and applications."[212]

In India, for example, Facebook joined forces with an Indian telecom operator to establish, in 2015, the Internet.org initiative (later renamed "Free Basics"), which offered free of charge access to select services and online content.[213] The concept is built on the fact that many people in the Global South are not in the position to pay the fees to obtain mobile data and therefore cannot access the Internet with their devices, so zero-rating allows them to do so. However, in implementing this strategy – which allowed Facebook to obtain millions of new members under the guise of "Internet.org"[214] – the social media company was accused of aiming to become indispensable for Indian users. It has been pointed out that Facebook's promotion of the idea of NN and its aim to have a "global connectivity platform based on zero-rating" are contradictory and cannot co-exist in practice.[215]

The issues that arise with zero-rating have major implications for elections, as this strategy creates an even stronger monopoly for massive content providers such as Facebook. It means that large corporations are the sole players with the financial capacity to offer their services at zero rates. Hence, they can determine what content the population can access depending on the countries. Zero-rating initiatives coupled with private messaging chatbots can be very problematic when a large number of the population can access social media platforms very quickly and for free but cannot access other websites. The most severe problem regarding elections in this context is that the electorate cannot easily fact-check disinformation distributed at large via social media platforms and private messaging apps.[216]

---

[210] See: https://unesdoc.unesco.org/ark:/48223/pf0000231162, p. 78, based on Barbara van Schewick, 6 May 2014, ***The Case for Rebooting the Network-Neutrality Debate***, The Atlantic. Available at: http://www.theatlantic.com/technology/archive/2014/05/the-case-for-rebooting-the-networkneutrality-debate/361809/.

[211] A. Daly, 2016, Net neutrality in Australia: The debate continues, but no policy In sight. In ***Net Neutrality Compendium***, Springer.

[212] V. K. Singh, 2015, Permit zero-rating schemes for a limited period, p. 1. Available at: https://perma.cc/4F7T-F87P.

[213] Carrillo, 2016.

[214] D. Banis, 2019, ***How 'Zero-Rating' Offers Threaten Net-Neutrality In The Developing World***. Available at: https://www.forbes.com/sites/davidebanis/2019/02/18/how-zero-rating-offers-threaten-net-neutrality-in-the-developing-world/#77ea286a3b41.

[215] A. J. Carrillo, 2015, Having Your Cake and Eating It Too: Zero-Rating, Net Neutrality, and International Law, ***Stan. Tech. L. Rev., 19***, p. 369.

[216] See: https://www.delianproject.org/_files/ugd/f769a5_d5f5910c553a4281b372560cb4bada5d.pdf

An example of the combination of zero-rating, automated messages, and disinformation was the 2018 Brazilian election. Zero-rating laid the ground for millions of people to access to message contents for free. But, in this context, tailored mass messages were used for political purposes, reaching users without their consent and circumventing WhatsApp span controls. These messages disseminated disinformation and misinformation, while users might not always had access to alternative content that could allow them to verify the integrity of the information received. As messages containing disinformation and misinformation often came from family members or friends through WhatsApp chats, users were less likely to contest them. They were therefore effective in contributing to shaping the elections' outcome.

4.7. ELECTORAL VIOLENCE & GENDER IN ONLINE SPACES

Violence against women (VAW) remains one of the most serious obstacles to the realization of women's political rights today. It can virtually disenfranchise women in elections, with effects on society that multiply from the resulting democratic deficit. This is emerging as a concern for policymakers and practitioners across the political spectrum. Until recently, lack of data and the stigma attached to gender-based violence in many societies have kept violence against women in elections (VAWIE) on the margins of study.[217] Yet it is a barrier to women that exists in every country, with cumulative and intersecting layers of discrimination on the basis of multiple characteristics and identity factors such as gender identity and expression, sexual orientation, race, age, social and economic status, disability, education level, occupation, marital status, religion, ethnicity, national origin or status, and so forth.

According to the "Violence Against Women in Elections Online: A Social Media Analysis Tool" developed by IFES, gender-based online violence (VAWIE-Online) is an umbrella term that captures a broad range of abusive, harassing, degrading and violent discourse circulating on the Internet or mobile technology across a range of intensities, from sexist slurs to direct threats of physical harm.[218]

Women frequently cite the threat of widespread, rapid public attacks on personal dignity as a factor deterring them from entering politics.[219]

Although it may comprise physical, sexual, or economic acts of aggression, electoral violence most often takes the form of psychological attacks. Socio-psychological violence is by far the most pervasive form of electoral violence experienced by women and the most widespread form of online violence. Indeed, in sample data, the proportion of intimidation and psychological acts of violence experienced by women was nearly three times the same proportion among men (a ratio of 28:10). Psychological violence is an "informal means of control [that] includes systematic ridicule, ostracism, shame, sarcasm, criticism, disapproval, exclusion and discrimination".[220]

Coupled with threats of physical and sexual violence, these forms of violence degrade, demoralize, and shame the individuals at which they are targeted. Online violence and abuse against women create a hostile environment with the aim of shaming, intimidating or degrading women.[221] Not all forms are crimes, but all impact the human rights of women. In a recent poll commissioned by

---

[217] J. Ballington, G. Bardall and G. Borovsky, ***Preventing violence against women in elections: A programming guide***, UN Women and UNDP, 2017. Available at: https://www.unwomen.org/en/digital-library/publications/2017/11/preventing-violence-against-women-in-elections.

[218] ***Violence Against Women in Elections Online: A Social Media Analysis Tool***, IFES, 2019. Available at: https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf.

[219] G. Bardall, ***Breaking the Mold: Understanding Gender and Electoral Violence***, IFES, December 13, 2011. Available at: https://www.ifes.org/publications/breaking-mold-understanding-gender-and-electoral-violence.

[220] Ibid., p. 8.

[221] S. Pinto, What is online violence and abuse against women?, ***Amnesty International***. Available at: https://www.amnesty.org/en/latest/campaigns/2017/11/what-is-online-violence-and-abuse-against-women.

Amnesty International in eight countries, nearly a quarter of women surveyed had experienced online abuse or harassment. A report released by the United Nations Broadband Commission called violence against women online a "problem of pandemic proportion."[222]

The report found that 73 % of women online have been exposed to or experienced some type of cyberviolence. Among the 86 countries included in the survey for the report, only 26 percent of law enforcement agencies have taken action against such violence. ICTs —especially social media—are frequently used as tools of gender-specific electoral and political violence.[223] There is overwhelming, global evidence of ICTs being used to perpetrate a broad range of violent acts against women during elections and in public life, especially acts that inflict fear and psychological harm.[224] ICTs may be used directly as a tool of intimidation by threatening or inciting physical violence against women candidates, voters, or representatives. Such cyber harassment or intimidation includes sending abusive, threatening or obscene emails, explicit threats of physical and/or sexual violence and encouraging strangers to physically harm the targeted person, which in some cases results in actual physical assault.[225]

Acts of VAWIE-Online may also involve spreading reputation-harming lies, electronic sabotage in the form of extensive spam and damaging viruses, impersonating the targeted person online and sending abusive emails or fraudulent spam, blog posts, tweets, and other online communications in the survivor's name or subscribing survivors to unwanted email lists, resulting in hundreds of unwanted messages daily. Such attacks can be perpetrated by both strangers and individuals known to the survivor, as well as by proxy stalkers and "cyber-mobs."

4.8. VIOLENCE AGAINST JOURNALISTS

While the Internet and social media have opened new opportunities to share information and ideas, those who felt their power was threatened by the increased reach of critical voices and independent sources of news have also enhanced their efforts to silence them. In some instances, this has been done by censoring specific pages or users or blocking entire websites. Yet, in others, it has entailed shutting off access to the Internet for a whole community, city, or country. With time, the attempts led by actors aiming to curtail freedom of expression gained sophistication using new technologies, by resorting for example to automated social media accounts or to selective bandwidth throttling. Tactics have also included the manipulation of information flows to diminish the visibility of journalistic work, and actions to get journalists to disseminate unverified content. These strategies may be less visible, yet nonetheless remain a fundamental threat to free, independent, and pluralistic media and, therefore, to democracy.[226]

As the UNESCO's World Trends Report 2017-2018 states, "there is no media freedom without safety, nor can there be independence or pluralism, when journalists work in fear."[227] At the same time, attacks against those performing journalistic functions – also including bloggers and citizen

---

[222] Combatting Online Violence Against Women & Girls: A Worldwide Wake-up Call, ***Broadband Commission Working Group on Gender***, September 2015. Available at: https://en.unesco.org/sites/default/files/highlightdocumentenglish.pdf.

[223] See: https://memo98.sk/uploads/content_galleries/source/memo/fiji/how-women-politicians-on-fiji-treated-on-face-book.pdf.

[224] See: https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online.

[225] ***Violence Against Women in Elections Online: A Social Media Analysis Tool***, IFES, 2019. Available at: https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf.

[226] D. Arnaudo, ***A New Wave of Censorship: Distributed Attacks on Expression and Press Freedom***, https://www.cima.ned.org/wp-content/uploads/2018/05/CIMA_A-New-Wave-of-Censorship_web_150ppi.pdf.

[227] UNESCO, World Trends in Freedom of Expression and Media Development, Global Report 2017/2018. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065.

journalists – have increased in the past decades, and impunity regarding these crimes prevails.

The dangers currently faced by these actors have serious implications for elections, as was discussed at length during UNESCO's 2019 World Press Freedom Day conference (held under the theme "Media for Democracy: Journalism and Elections in Times of Disinformation") and reflected in the Addis Ababa Declaration that emerged from the event.[228] There are increasingly digitally-intensified patterns of threats and violence against journalists and other actors who contribute to public debate, including specific and gender-based threats and violence against women journalists.[229] Deep investigation, including using big data analysis, of attacks on women journalists reveals a range of gender-specific tactics.[230] Some of the key forms of threats and attacks faced by journalists, bloggers and others contributing to journalism content include illegitimate surveillance, tracking, hacking and doxing (publishing personal information about people without their consent), fake domain attacks,[231] phishing,[232] online harassment and distributed denial-of-service (DDoS) attacks.[233]

A worrying trend is some governments' use of laws that are worded in a vague manner to justify disproportionate online censorship and the surveillance of journalists.[234] The use and abuse of surveillance tools and their impact on human rights has been an area of growing concern. In May 2019, for instance, the UN Special Rapporteur on Freedom of Opinion and Expression, issued a Special Report on "Surveillance and Human Rights"[235] which calls for a moratorium on the sale of surveillance tools to States, until adequate legal safeguards for freedom of expression can be put in place.

The published UNESCO-commissioned survey on online violence against women journalists showed that, while 25% of the surveyed women journalists said that they had reported online attacks to their employers, the most common response was no action being taken (10%), followed by advice of the type "to grow a thicker skin" or "toughen up" (9%). In 2% of the cases, employers asked them what they had done to provoke the online violence they received. Moreover, the survey revealed women journalists' low level of access to systems and support mechanisms to deal with online violence, as well as their insufficient awareness regarding measures, policies, and guidelines to tackle the problem.

---

[228] UNESCO, Addis Ababa Declaration, https://en.unesco.org/sites/default/files/wpfdaddisdecl3_may.pdf.

[229] See: T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, https://unesdoc.unesco.org/ark:/48223/pf0000371486, p. 8.

[230] J. Posetti, N. Shabbir, D. Maynard, K. Bontcheva and N. Aboulez, ***The Chilling: Global trends in online violence against women journalists***, UNESCO, 2021. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000377223.

[231] Digital attacks on journalists can take the guise of a fake domain (website) that silently collects information that a journalist enters on the site thinking that it is legitimate.

[232] Phishing (or "Password-Stealing Phishing") takes place when malicious actors gain access to someone's information through the creation of a website that imitates the login prompt of an online service. The victim is thus lured into entering her/his username and password, which are transmitted to those behind the attack.

[233] A distributed denial-of-service (DDoS) attack aims to disrupt the normal traffic of a server, service or network. It does so by overwhelming the target or the infrastructure that surrounds it with a flood of Internet traffic, which prevents regular traffic from arriving to it.

[234] S. Waters, The Effects of Mass Surveillance on Journalists' Relations With Confidential Sources, ***Digital Journalism***, Vol. 6, Issue 10, p. 1294, 2018.

[235] Report of the Special Rapporteur on Freedom of Opinion and Expression, A/HRC/41/35, 28 May 2019, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/148/76/PDF/G1914876.pdf?OpenElement.

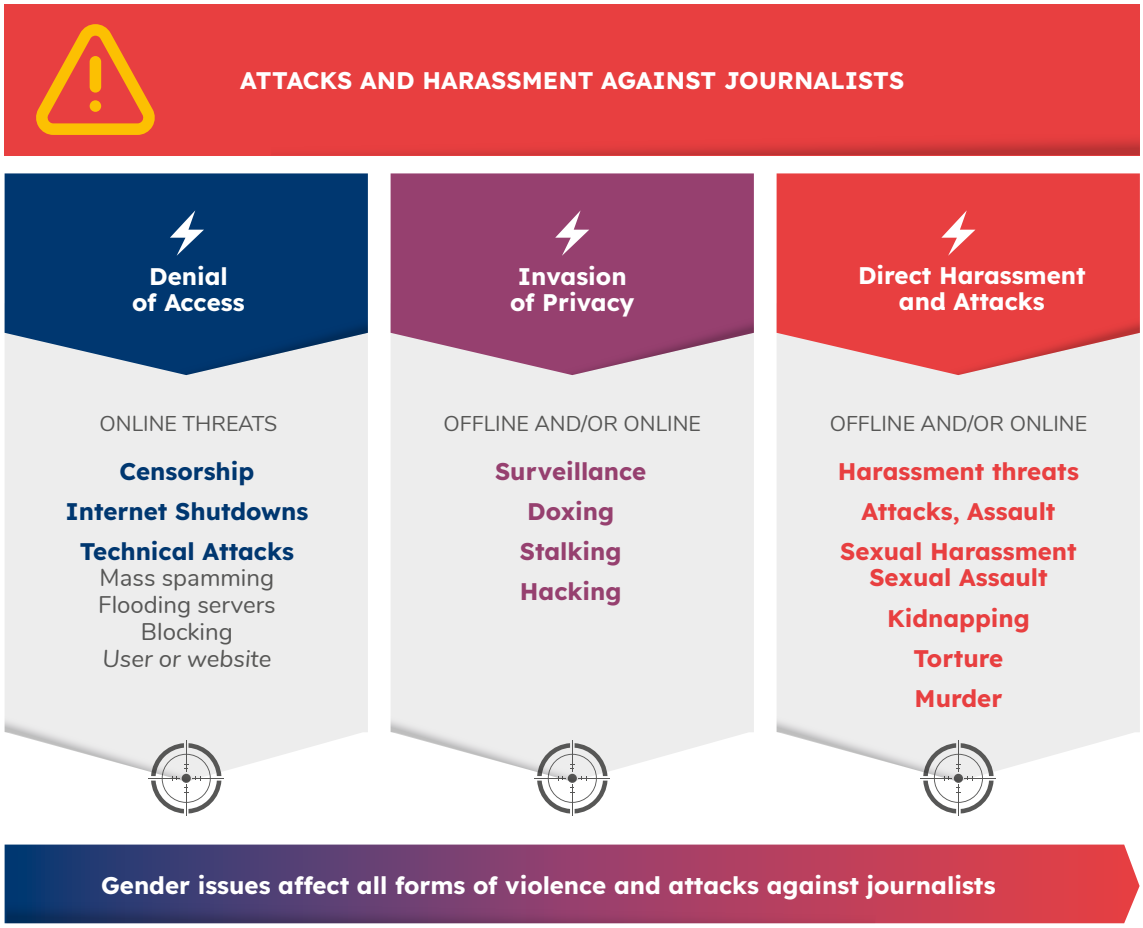


FIGURE 7: EXAMPLES OF ATTACKS AND HARASSMENT AGAINST JOURNALISTS[236]

These findings highlight the urgent need for responses in this area, including enhanced legal protection, employers' provision of online safety support and training for women journalists that are part of their staff, and news organizations' adoption of gender-sensitive procedures and systems for the identification, reporting and monitoring of online violence against their personnel. It is also recommended for Internet companies to include in their transparency reports details about the types of reports they receive and the actions they consequently take. Multi-stakeholder collaborative efforts – involving CSOs, journalists, networks, and researchers – would also serve to advance understanding and develop evidence-based responses and support for women journalists.[237]

[236] S. McCabe, T. Chorbacher, M. Churchill and E. Kirkland, ***Intensified Attacks, New Defences. Developments in the Fight to Protect Journalists and End Impunity***, UNESCO, 2019, p. 45. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000371487?fbclid=IwAR2FAIrs5INt5ibUb_gHxfNYaFEzTYITbMeEob8ZjXfjbDFAII1ad8vHAgk; based on an idea by IWMF and Trollbusters, 2018, p. 22, https://www.iwmf.org/wp-content/uploads/2018/09/Attacks-and-Harassment.pdf.

[237] See: https://unesdoc.unesco.org/ark:/48223/pf0000375136.

ACTIVITY IV

The following activity has the objective of determining the participant's level of knowledge on the impact of social media platforms on citizens during electoral processes, the way that algorithms can amplify misleading content, the concepts of voter suppression, Internet shutdown and disruption of Net neutrality via zero-rating, the critical role of cybersecurity, digital campaigning and how technology contributes to increasing violence against women and journalists during elections.

Suggested guiding questions for a discussion:

I. Why do social media platforms enhance the opportunities for citizens to retrieve information that is important for their voting decisions? What is their impact in places where freedom of expression and access to information are restricted?

II. Recent electoral processes showed that social media networks can also be used for harmful purposes that. Can you mention specific ways in which algorithms amplify engagement of misleading content that undermine the integrity of elections all along the electoral cycle?

III. What is voter suppression? What is an Internet shutdown? How do they affect freedom of expression?

IV. Please explain the concept of "disruption of Net Neutrality via zero-rating"

V. Why does effective cybersecurity play a critical role in the EMBs operational planning? Elections rely on varying combinations of manual and technology-based procedures. How can a breach of cybersecurity affect elections?

VI. Can you describe the benefits and dangers of digital campaigning? What is micro-targeting?

VII. Why does violence against women in elections (VAWIE) remain one of the most serious obstacles to the realization of women's political rights today? Please explain.

VIII. Can online attacks and harassment against journalists during the electoral cycle affect freedom of expression?

IX. A worrying trend is some governments' use of laws that are worded in a vague manner to justify disproportionate online censorship and the surveillance of journalists. Can you provide any examples?

5. TACKLING DISINFORMATION ALL ALONG THE ELECTORAL CYCLE

OBJECTIVES OF THIS SECTION

- Identify the measures that can be taken to tackle disinformation along the electoral cycle and in the short, medium and long term.
- Examine media regulation during elections and regulation, self-regulation and co-regulation of online content.
- Provide an overview of codes of practice by Internet intermediaries.
- Understand the relevance of voter education and media information literacy.
- Understand the importance of building capacities among judicial actors and addressing (online) violence against women and other vulnarable groups.

With the advances in technology in recent years, electoral management has increased in complexity in ways that often cannot be tackled without the help of ICTs. EMBs have much to gain from upgrading their capacities to harness the use of technology in diverse ways. In times of the COVID-19 pandemic and decreasing voter turnout, EMBs face ever-rising pressure – from policymakers and citizens alike – to offer public services online. EMBs thus need to explore ways to make processes more accessible for voters, candidates, parties, elected representatives, media, and the general public.[238]

These efforts can include offering multiple voting channels to choose from, which differ from traditional election-day voting in polling stations either in terms of time (e.g., advanced voting), space (e.g., home voting, voting in public spaces such as supermarkets, county centers, embassies, etc.),[239] or medium (e.g., voting on paper or electronically through ballot scanners, electronic voting machines, or Internet voting).[240] EMBs need to address the fact that, with a more mobile population, more changes come, and strengthening ICT-related capacities becomes mandatory.

Also, because disinformation, misinformation and mal-information are multi-faceted problems, they require a range of responses from multiple actors, not only the EMBs but also from the governments, media, Internet intermediaries, political actors, CSOs, academia, and individual citizens. Measures against online disinformation should take account of the different types of potential harm caused by different types of disinformation all along the electoral cycle. Measures can have varying levels of impact and be designed for short-, medium- and long-term periods. The ways to tackle disinformation can be grouped in terms of four categories[241]: preventing measures, identification and monitoring measures, regulatory and non-regulatory measures, containing or corrective measures.

---

[238] R. L. Pintor et al., 2002.

[239] R. Krimmer, D. Duenas-Cid and I. Krivonosova, 2021, New methodology for calculating cost-efficiency of different ways of voting: is internet voting cheaper?. ***Public Money & Management***, 41:1, DOI: https://doi.org/10.1080/09540962.2020.1766222.

[240] OSCE/ODIHR, 2013; R. Krimmer and M. Volkamer, 2005, Bits or Paper? Comparing Remote Electronic Voting to Postal Voting, In ***EGOV (Workshops and Posters)***.

[241] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

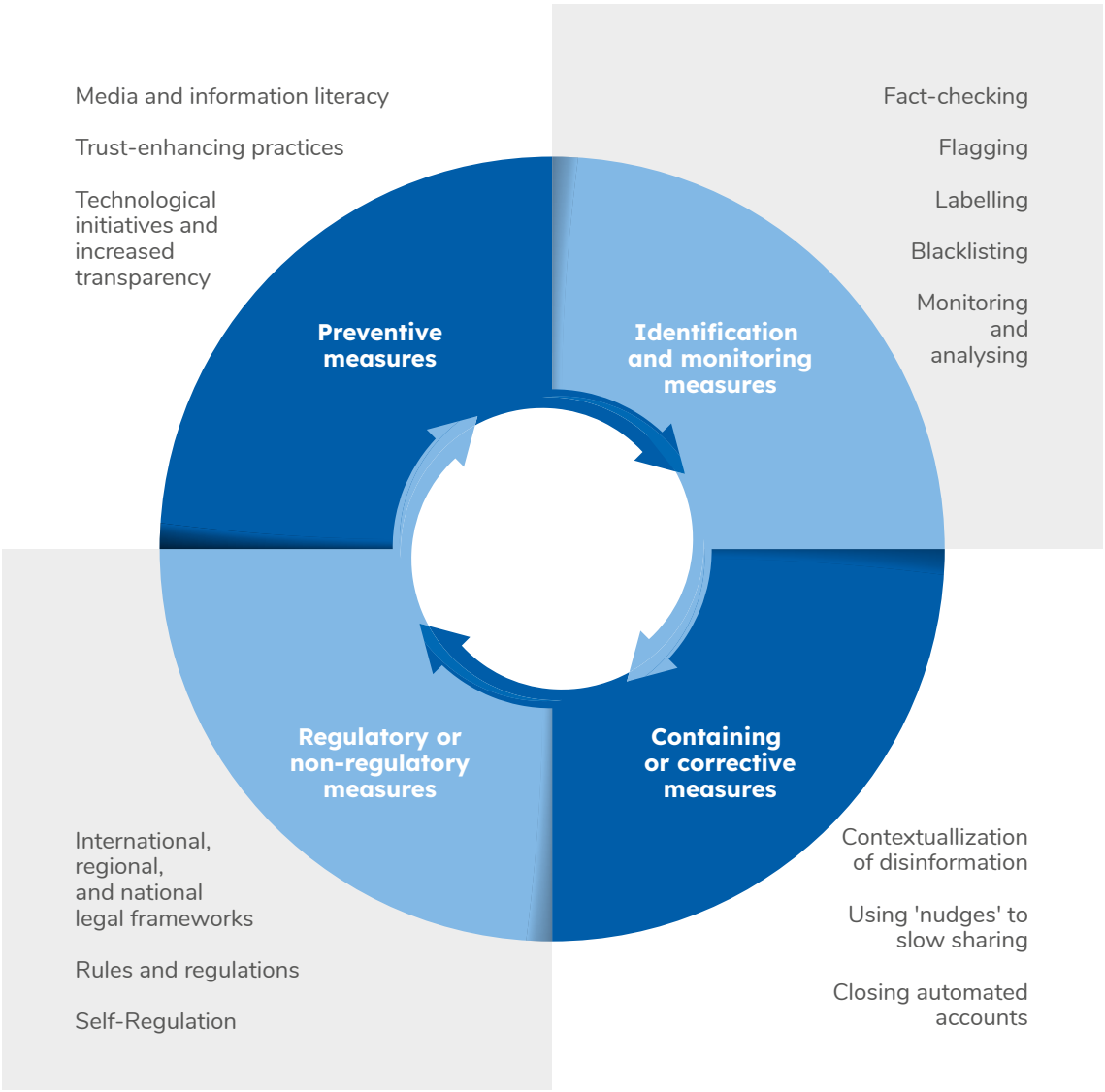


FIGURE 8: WAYS TO TACKLE DISINFORMATION

At the same time, EMBs must also consider all the problems created or amplified by technology, like inflammatory speech. In all cases, EMBs need to contemplate the digital dimension and have adequately planned and budgeted staff and resources throughout the electoral cycle without neglecting any aspect.

5.1. REGULATION, SELF-REGULATION, AND CO-REGULATION OF ONLINE CONTENT

The issue of the regulation of online content during electoral periods entails multiple complexities.[242] The global nature of the Internet makes attempts to regulate it difficult. Illegal or harmful online content can be created and disseminated much faster and on a broader scale than offline expressions. In addition, a relatively reduced number of companies dominate the online discourse regulated through their own global terms of service and community standards – rather than laws aligned to international human rights standards, but instead following a commercial logic.

Private companies make internal content regulation decisions that are often automated, and entail limited human review, factors that challenge traditional practices of norm formation and enforcement. More and more countries are trying to enforce certain restrictions during electoral processes considered necessary to ensure free and fair elections, which are compatible with Article 19 of the ICCPR if they are proportional and non-discriminatory. As stated in the report by the UN Special Rapporteur on Freedom of Opinion and Expression Frank La Rue in 2014, "the adequate regulation of political communications is crucial to ensure a just and equitable space for public dialogue and access to information. In a democratic society, elections must never be ruled by the market logic with those having greater access to financial support controlling the public debate through their disproportional access to publicity and media."[243]

Media regulations during elections are implemented to ensure, among other aims, that the electorate can access a range of different perspectives and opinions and that candidates compete on a "level playing field" and have the right to reply to claims against them. Broadcast media, which receive a license to disseminate information through airwaves that are part of a limited spectrum considered a public good, are particularly subject to conditions set to promote fairness and balance in electoral coverage.

However, most legislation and rules governing elections and related media self-regulatory tools do not always apply to digital platforms, social media, and social messaging. A clear example relates to the "silence periods" or "campaign moratoriums" that many countries enforce immediately before election day. There have been cases in which, even if a government prohibits the publication of polling results right before an election, online news outlets, blogs, etc. – particularly those located in other countries – have made some of these results public earlier than the closing of all polls, allowing citizens from the State where elections were held to access them.[244] Social media has also been used to disseminate elections results before official announcements and to circulate information during campaign moratoriums.[245] Likewise, political advertising often spreads online and via social messaging even during the imposed silence periods.

While the Internet can help to level the playing field by enabling candidates with fewer resources to share their messages and mobilize voters at lower costs, social media platforms tend to benefit candidates that dispose of more financial resources. Political parties, political groups, and candidates who have more money can disproportionately benefit from micro-targeting. Data mining and data harvesting have the possibilities for circumventing political advertising rules, whether about spending ceilings and disclosure or truthfulness in political advertising, primarily

---

[242] See: B. Martin-Rozumilowicz and R. Kužel, ***Social Media, Disinformation and Electoral Integrity***, IFES working paper, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.

[243] See: https://undocs.org/A/HRC/26/30, para. 77.

[244] See: A. Puddephatt, 2019, ***Social media and elections***, UNESCO, Paris. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000370634.

[245] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***.

due to their transnational character. In turn, although online media and social platforms are increasingly functioning as information gatekeepers, they are neither bound by the same rules nor by ethical standards for news reporting.

Adding complexity to these matters is that social media platforms are often protected from liability in many jurisdictions. They are considered primarily aggregators or carriers of content produced by others – rather than publishers – and therefore hold no editorial responsibility.[246] As social media companies' impact on democracies has become more evident, the notion of their total exemption from liability is being contested.[247]

Indeed, much of the debate on the regulation of social media networks focuses on whether these networks are considered merely OTTs or editors, given their considerable control over the dissemination of user-generated content. Some experts support the view that social networks are media outlets, and as such, they should be regulated by a statutory legal framework. Others argue that government regulation of such platforms would be detrimental to freedom of expression and challenging to enforce.[248] They call for caution regarding attempts to treat social media platforms as media outlets and regulate them similarly to publishers, given that they fulfil different functions – including hosting, online distribution, and, only in some limited circumstances, editing or commissioning content over which they hold editorial control.[249]

5.1.1. LIABILITY OF OVER THE TOP SERVICE PROVIDERS (OTTS)

When examining approaches to online content regulation, the issue of liability of Over the Top service providers (OTTs) emerges as a critical topic. In the European Union, the United States, and many countries around the world, Internet Service Providers are exempted from liability for the content that they are simply hosting or organizing (not creating), and which was placed on their website or social media platform by third parties.[250] By contrast, other countries apply to OTTs the ordinary rules of civil and criminal liability.

The adoption of a liability exemption assumes that it would be bordering on the impossible for OTTs to monitor every piece of information uploaded by users – although it could be claimed, against this argument, that currently the main platform providers do monitor and manage users' content. This protection is essential, considering the role that OTTs play in allowing access to information and in the realization of freedom of expression, which also places them in a position of vulnerability vis-à-vis political pressure.[251]

---

[246] See: A. Puddephatt, 2019, ***Social media and elections***.

[247] Even though they are not necessarily legally binding, Meta, Google, TikTok and Twitter, have committed to tackling online abuse and improving women's safety on their platforms at the UN Generation Equality Forum in Paris in 2021. Companies have committed to offering more granular settings (for example, who can see, share, comment, or reply to posts), using accessible language throughout the user experience, providing easy navigation and access to safety tools, and reducing the reporting burden on women by proactively reducing the amount of abuse that occurs. See: https://forum.generationequality.org/.

[248] UNESCO, World Trends in Freedom of Expression and Media Development, Global Report 2017/2018. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065, https://en.unesco.org/world-media-trends.

[249] ARTICLE 19, 2018, https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf.

[250] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), ***Official Journal L 178***, 17/07/2000 P. 0001 – 0016, Article 14 and 15. In the United States: Telecommunications Act of 1996, Pub.L.No 104-104, 110 Stat, 56as amended by 47 U.S.C) § 230 (2000) - which "gives Internet intermediaries a privilege against certain lawsuits based on content provided by third parties" in J. M. Balkin, The Future of Free Expression in a Digital Age, Pepperdine Law Review, Vol. 36 N, 2008, p. 108; U. Kohl, Google: The Rise and Rise of Online Intermediaries in the Governance of the Internet and Beyond (Part 2), ***International Journal of Law and Information Technology***, Vol. 21, No. 2, 2013, p. 191.

[251] The UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression in the Report of 16 May, 2011 (A/HRC/17/27) in para. 43 states that "[...] censorship measures should never be delegated

Although aiming to protect OTTs, the "notice-and-takedown" regime of conditional liability has come under criticism, as it results in content being removed by OTTs based on notification, rather than on an objective assessment of the content of the message[252] and careful consideration of competing interests and defences.[253] Also concerning are problematic aspects derived from automated decision-making on content, through the use of algorithms,[254] and the influence of advertising on the information that can be accessed (or not) through social media platforms.[255]

Moreover, also worrying are private companies' shutdown of certain fora where users could openly discuss issues and exchange ideas or, in fact, mobilize, due to fear of legal and financial consequences. Other issues that have been pointed out as problematic concern the limited room for recourse that a user has to challenge a takedown, the fact that OTTs tend to over-censor to avoid being held liable, and the lack of transparency in OTTs' decision-making processes.[256] In some instances, notice-and-takedown procedures have been found to be subject to abuse both by governments and private actors, resulting in excessive liability of OTTs and restrictions on legitimate forms of expression.[257]

BOX 10: EUROPEAN COURT JURISPRUDENCE

In the European Union, the exemption of liability of Internet intermediaries derives from the European Union E-Commerce Directive,[258] and is conditional upon the intermediary acting expeditiously to remove unlawful content once it is notified about it, although the directive does not require that intermediaries monitor the content posted by users – that is, they are not obliged to look for illegal content actively.[259]

The European Court jurisprudence on social media platforms has been evolving in recent years. In ***Delfi v. Estonia***, in 2015, the European Court of Human Rights ruled that holding the news site Delfi liable for anonymous defamatory comments posted on

to a private entity, and [...]no one should be held liable for content on the Internet of which they are not the author. Indeed, no State should use of force intermediaries to undertake censorship on its behalf...". See also: J. Rosen, The Delete Squad – Google, Twitter, Facebook and the new global battle over the future of free speech, ***The New Republic***, April 29, 2013. Available at: https://newrepublic.com/article/113045/free-speech-Internet-silicon-valley-making-rules.

[252] Electronic Frontier Foundation, https://www.eff.org/free-speech-weak-link and http://www.michaelgeist.ca/2017/02/bogus-claims-google-submission-points-to-massive-fraud-in-search-index-takedown-notices/.

[253] See: Report of Special Rapporteur on Freedom of Expression and Opinion, 7 September 2012 (A/67/357), https://undocs.org/A/67/357.

[254] A difficulty posed by the use of automated systems is that speech may be incorrectly categorized due to the reinforcement training of the system responsible for removal of information. For instance, the word "Nazi" might be removed regardless of the context in which it was used, with the system being unable to provide a justification for why it arrived at a certain conclusion. This happened with the case of the "Napalm Girl" photo, which was taken down from Facebook as a result of nudity.

[255] S. Hill, Empire and the megamachine: comparing two controversies over social media content, ***Internet Policy Review***, 8(1), 2019.

[256] See: https://undocs.org/A/66/290.

[257] See: https://undocs.org/A/66/290.

[258] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), ***Official Journal L 178*** , 17/07/2000 P. 0001 – 0016.

[259] See: https://www.mediadefence.org/wp-content/uploads/2020/06/20140606-Delfi-intervention-FINAL-1.pdf and B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019.pdf.

it did not constitute a violation of freedom of expression. The Court cited, among other reasons for this interpretation, that the remarks included unlawful speech amounting to hate speech and that the commercial news website had not taken sufficient measures to filter them out and remove them fast enough.[260] This ruling and the one related to the ***MTE v. Hungary*** case[261] established the concept of 'notify and take-down' and were the first ones to be brought against OTTs in Europe. In another case in 2017, however, the Court highlighted that holding an Internet portal liable for third-party comments may have a chilling effect on freedom of expression via the Internet.[262]

In turn, the European Court of Justice[263] has repeatedly interpreted the liability exemption to apply only to those intermediaries "which are of a mere technical, automatic and passive nature"[264] and has recently, ruled, in the case of ***Eva Piesczek v. Facebook*** (2019), that in order to benefit from a limitation of liability, an OTT, "upon obtaining actual knowledge or awareness of illegal activities has to act expeditiously to remove or to disable access to the information concerned".[265] Furthermore, in 2018, the Council of Europe Committee of Ministers adopted a recommendation on the roles and responsibilities of Internet intermediaries,[266] calling for a "graduated approach"[267] to the liability for content, not departing significantly from the case of ***Delfi v. Estonia*** and leaving the ultimate responsibility with the States.

Thus, the European Union liability exemption principle has not proven as robust and impermeable as in the United States, where the courts have recognized[268] that the intention of US legislators in providing an exemption from liability[269] was to preserve the freedom of the Internet by protecting the providers from tort liability. The EU Digital Services Act tackles the roll-back of liability exemption, including the premises under which such liability is either limited or not.

---

[260] ***Delfi AS v. Estonia*** (Application No.64569/09), Judgement of the European Court of Human Rights of 16 June 2015.

[261] ***MTE v. Hungary*** (Application no. 22947/13), Judgement of the European Court of Human Rights of 2 February, 2016, Final 02/05/2016.

[262] ***Pihl v. Sweden***: https://hudoc.echr.coe.int/eng#{%22fulltext%22:[%22\%22PIHL%20v.%20SWEDEN\%22%22],%22itemid%22:[%22001-172145%22]}, cited in B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019.pdf.

[263] ***L'Oreal SA and others v. Ebay International A.G. and others***, Case C324/09, Judgement of the Court (Grand Chamber) of 12 July 2011, para. 119 where the provider has confined itself **"to a merely technical and automatic processing of data"** and para. 123 where the exemption is relevant as **"applying to the operator of an online marketplace where that operator has not played an active role allowing it to have knowledge or control of the data stored"**.

[264] E. Coche, Privatized enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online, ***Internet Policy Review***, 7(4), 2018.

[265] ***Eva Glawischnig-Piesczek v. Facebook Ireland Limited***, Case C18/18, 3 October 2019, para. 46.

[266] CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of Internet intermediaries.

[267] Ibid., para. 12 of the Annex.

[268] ***Carafano v. Metrosplash.com Inc***, US Court of Appeals for the Ninth Circuit, No 02-55658, 13 August 2003.

[269] 47 U.S. Code § 230 - Protection for private blocking and screening of offensive material.

5.1.2. REGULATION

The regulation of online content has become the focus of intense debate. In recent years, approximately 52 countries worldwide have adopted legislation against disinformation, and others have also introduced restrictions on online content and social media through other special laws.[270]

IFES' Working Paper on "Social Media, Disinformation and Electoral Integrity" (2019) presents an overview of existing regulatory and legal approaches to disinformation, categorizing them as follows:

1. Regulation of content through blocking or removal; criminalization of information dissemination; criminalization on the grounds of defamation or hate speech.
2. Regulatory or legal mandates to monitor social media.
3. Classifying social media as traditional media.
4. Data privacy laws.
5. Political finance regulation that pertains to social media.
6. Regulation of dissemination methods.
7. Regulation of specific technology companies or social media platforms.
8. Voter education mandates.[271]

It is yet too early to determine whether regulatory initiatives' outcomes are positive and effective, but they have been the target of much criticism. The international NGO ARTICLE 19 has, for instance, observed that «many of the recent legislative initiatives related to the Internet and social media companies tend to give disproportionate censorship powers to the State, whether through prison terms, fines or content blocking powers, chilling free expression, or to outsource regulation to private companies with no proper integration of international standards».[272] The regulatory entities linked to these initiatives are generally not fully independent, and the laws do not always allow for an appeal or judicial review of their decisions.[273]

Laws that attempt to combat disinformation often include provisions forcing big technology companies to remove or block content flagged by third parties quickly. In some cases, failure to comply with such notifications can mean civil liability and fines for Internet intermediaries. As such, the companies are under pressure to remove content regardless of whether it is legal or not in order to avoid the fines, rather than meeting international law standards.

There could be a risk that opposition candidates, for instance, will not have tools for communication or mobilization during elections, especially in countries without proper oversight mechanisms if government entities or private companies determine what gets taken down.

Besides, while some governments' efforts are being implemented in good faith to tackle disinformation and hate speech, others appear to be aiming at increasing control over speech for political gain[274] – silencing human rights activists and journalists under the pretext of "fake

---

[270] See: B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019.pdf. The last category, focused on voters' education mandates, is addressed separately in this Chapter, on Section 3.2.3.

[271] Ibid.

[272] See: https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

[273] See: B. Martin-Rozumilowicz and R. Kužel, 2019. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019.pdf.

[274] K. Bontcheva and J. Posetti (eds.), 2020, ***Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression. Broadband Commission research report in 'Freedom of Expression and Addressing Disinformation on the Internet'***. Available at: https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf.

news" by having their accounts suspended and their content taken down. Even in the case of well-intentioned laws, some have also proven to be difficult to enforce in practice, given the tensions between the international operations of social media platforms and States' jurisdiction being limited to their territory. In addition, it is hard for any regulatory approach to keep up with the pace of technological developments.

An analysis of more than twenty recent legal initiatives to tackle different types of online harm, undertaken by the multi-stakeholder Global Network Initiative (GNI), is in line with the considerations above, referring to "vague or broad definitions for the content and/or companies covered by the scope of regulations; deputizing private companies as judge, jury, and executioner of the legality of user content and conduct; overreliance on automated moderation tools; and potential privacy infringements, including by prohibiting or undermining encryption".[275]

BOX 11: GERMANY'S NETWORK ENFORCEMENT ACT (NETZDG)

Often cited among the examples of recent legislation is Germany's Network Enforcement Act (NetzDG), which was passed in 2017 to combat hate and extremist content online. The law requires social media networks to establish clear procedures for flagging content and managing complaints, and to block or remove content that violates restrictions on hate and defamatory speech in the German Criminal Code.[276] The social media networks that repeatedly fail to comply with the NetzDG may be fined up to fifty million euros.

What makes content "manifestly" illegal is, in the first instance, left up to human or algorithmic judgment. As a result, the NetzDG incentivizes intermediaries to remove demeaning content that could potentially violate the Criminal Code.[277] The obligation for platforms to remove manifestly unlawful content within 24 hours is especially challenging, both for the operators of social platforms and the country's legal system – as it would be difficult to achieve unless a judicial system reform also takes place.[278]

Despite describing it as "a good-faith effort to deal with widespread concern over online hate and its offline consequences", in 2019 the then UN Special Rapporteur on Freedom of Opinion and Expression, David Kaye, noted what he found to be shortcomings in terms of definitional vagueness and the significant fines imposed on companies in case of non-compliance.[279]

BOX 12: FRENCH LAW OF 2018 CONCERNING THE FIGHT AGAINST INFORMATION MANIPULATION

In relation to electoral processes, one recent effort was the adoption by the French Parliament, in 2018, of a law to counter information manipulation before and during elections. During the three months preceding an election, the law stipulates that an interim

[275] See: https://globalnetworkinitiative.org/content-regulation-policy-brief/.

[276] Platforms are obliged to remove or block access to "manifestly unlawful content" within 24 hours, "unlawful content" within 7 days and, if they receive over 100 complaints per year, they have to publish reports every six months on how they dealt with flagged content. K. Bontcheva and J. Posetti (eds.), 2020.

[277] Parliamentary Assembly of the Council of Europe (PACE) Report 15028, ***Democracy hacked? How to respond?***, 8 January 2020. Available at: http://assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=28319&lang=EN.

[278] Ibid.

[279] See: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, on hate speech online, 9 October 2019 (A/74/486), https ://undocs.org/A/74/486, para. 32.

judge can act "with appropriate and necessary measures" to stop the dissemination of disinformation. The ruling must be issued within 48 hours from the receipt of a complaint from authorities, electoral candidates, political groups, or individuals.

In this framework, disinformation is defined as "inaccurate or misleading accusations or allegations with the aim of changing the sincerity of a vote". The law also requires online platforms to cooperate by establishing mechanisms to flag content, ensuring algorithmic and advertising transparency, promoting mainstream media content, and implementing media and information literacy initiatives. It also gives additional power to the broadcasting regulator to monitor platforms and to revoke the licenses of foreign broadcasters found to be disseminating misinformation.[280]

BOX 13: AMENDMENTS TO THE BRAZILIAN ELECTORAL CODE

In Brazil, several bills proposing to criminalize disinformation during elections have been introduced to the Congress. A law amending the electoral code, adopted in September 2019, defines the crime of "slanderous denunciation for electoral purpose" and foresees a penalty of two to eight years of prison for it. Following a 2017 resolution, the electoral court can request platforms to remove content about candidates that is "known to be untrue".[281] A joint parliamentary inquiry committee was also set up to investigate how disinformation and profiling were deployed to influence elections results in 2018.[282]

A trend emerging with the COVID-19 pandemic has been the adoption of laws and regulatory measures that:

> "...effectively criminalised acts of producing or sharing information deemed to be false, misleading and/or contradicting official government communications about COVID-19. Emergency decrees giving political leaders sweeping new powers were among these measures, along with the application of existing emergency acts to COVID-19 disinformation to enable arrests, fines and jail time for associated offences..."[283]

In this context, the UN Special Rapporteur for Freedom of Opinion and Expression published a report[284] and OHCHR made available guidance[285] that reflected concerns about new restrictions of rights during the pandemic – including freedom of expression, access to information and privacy, among others.[286]

It is important to recall that any attempt to regulate online content and its dissemination should balance the rights to freedom of expression and access to information with the protection of

---

[280] K. Bontcheva and J. Posetti (eds.), 2020.

[281] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***.

[282] K. Bontcheva and J. Posetti (eds.), 2020.

[283] K. Bontcheva and J. Posetti (eds.), 2020, p. 1110.

[284] See: https://digitallibrary.un.org/record/3862160?ln=en#record-files-collapse-header.

[285] See: https://www.ohchr.org/EN/NewsEvents/Pages/COVID19Guidance.aspx.

[286] See: https://www.un.org/en/un-coronavirus-communications-team/we-are-all-together-human-rights-and-covid-19-response-and.

other civil and political rights, such as participation, privacy, and freedom from discrimination. Furthermore, as in the offline sphere, restrictions to online freedom of expression should meet the three-part test outlined in Article 19 of the ICCPR (it should be prescribed by law, pursue a legitimate aim, be necessary and proportionate). Regulatory responses to disinformation and hate speech should be evaluated according to these standards. In this regard, the Human Rights Committee General Comment 34 explains that:

> "Permissible restrictions generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3 [of Article 19]. It is also inconsistent with paragraph 3 to prohibit a site or an information dissemination system from publishing material solely on the basis that it may be critical of the government, or the political and social system espoused by the government".[287]

The UN Special Rapporteur on Freedom of Opinion and Expression has also recalled that "States should not use Internet companies as tools to limit expression that they would be precluded from limiting under international human rights law" and warned about the pressure for these companies to use automated tools that could lead to pre-publication censorship and disproportionate outcomes, given that these filters cannot detect subtleties in language, nor the different impact that specific content can have in other locations.[288]

In turn, the blocking of entire websites or social networks as a content regulation approach (often used during elections, protests, and other politically relevant occasions) has been described in a 2011 report by the UN Special Rapporteur on Freedom of Opinion and Expression as "an extreme measure – analogous to banning a newspaper or broadcaster – which can only be justified under international standards, for example, where necessary to protect children against sexual abuse".

5.1.3. SELF-REGULATION AND SOLO-REGULATION

Self-regulation is defined as "a mechanism of voluntary compliance at sector or industry level: legislation plays no role in enforcing the relevant standards. Its raison d'être is holding members of self-regulatory bodies accountable to the public, promoting knowledge within its membership, and developing and respecting ethical standards".[289]

An interesting initiative under this category has been the proposal, by the NGO ARTICLE 19, of creating social media councils, either at the national or international level, consisting of a multi-stakeholder mechanism of accountability for the moderation of content on social media. This model was endorsed in 2019 by then UN Rapporteur on Freedom of Opinion and Expression, David Kaye. It consists of an approach whereby social media platforms and other involved stakeholders join a mechanism that does not create legal obligations but relies instead on the voluntary compliance by platforms that commit to respect and implement the social media council's decisions.

Also generally included under the category of self-regulation are situations that some experts prefer to define as "Solo-regulation"[290] or "regulating speech by contract", in which a private company unilaterally controls content on its own platform, according to its own internal rules".[291]

---

[287] CCPR/C/GC/34 General Comment No. 34 on Article 19 of the ICCPR states that, "The scope of paragraph 2 embraces even expression that may be regarded as deeply offensive".

[288] See: https://undocs.org/A/74/486, 2019.

[289] See: https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

[290] Ibid., based on Marko Milosavljevic and Sally Brougthon Micova in their article 'Banning, Blocking and Boosting: Twitter's solo-regulation of expression', Medijske Studije / Media Studies, 2016, 7 (13), pp. 43-58.

[291] See: https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

This is the approach underlying the terms of service and community guidelines that users agree to when joining social media platforms, such as Facebook, Twitter and YouTube, in certain countries.

The often-cited shortcomings of the "Solo-regulation" approach have been the lack of transparency and insufficient protection of freedom of expression and other human rights. Companies can be vulnerable to political pressures to remove or filter specific content, with the consequent privatization of online censorship. As a result, the opacity of companies' decision-making processes could leave citizens without any mechanisms for protecting their freedom of speech.[292] The need for these companies' services as messengers and social media leaves the user with no choice but to accept their terms and conditions of use, with nothing concrete to mediate between the user and the service provider.

5.1.4. CO-REGULATION

"Co-regulation", which has also been referred to as "Regulated self-regulation", is a model that entails private regulation (either in the form of self-regulation or solo-regulation) that is supported or actively encouraged by the government,[293] and is typically underpinned by legislation.[294] The NGO ARTICLE 19 includes the case of Germany under the Network Enforcement Act (NetzDG) within this category, as it sets up a regulated self-regulatory agency for social media.[295]

A type of co-regulatory model that is emerging as a trend, including around electoral periods, is centred around the adoption of Codes of Conduct. These include for example the EU Code of Conduct on Hate Speech and the EU Code of Practice on Disinformation (for details see Section 5.3.1).

5.2. THE RESPONSE BY SOCIAL MEDIA PLATFORMS

As mentioned above, governments are struggling to develop effective public policies to address disinformation, misinformation and mal-information. The general acceptance of a principle of limited liability for OTTs has shifted to increasing calls for Internet intermediaries to play a more active role as information gatekeepers.[296] Considerations regarding the integrity of elections are central in calls for social platforms to take a more active role concerning emerging challenges.

Mainly, social media networks and other Internet intermediaries have focused on developing self-regulatory rules, usually in the form of a code of standards or terms of service on issues such as content removal and data processing practices. As part of their efforts to increase transparency and raise awareness of the increasing threats to freedom of expression online, Google, Facebook, Twitter, and other major Internet companies have also started making available transparency reports. These show the number of requests to take down content and access user data they received from governments and whether the company has complied.

---

[292] See: https://www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB-v2.pdf. and ARTICLE 19, 2018, https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf.

[293] See: https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

[294] See: https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf.

[295] Ibid., and https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

[296] UNESCO, World Trends in Freedom of Expression and Media Development", Global Report 2017/2018. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000261065.

5.3. CODES OF PRACTICE AGREED UPON BY INTERNET INTERMEDIARIES

5.3.1. EU CODE OF PRACTICE ON DISINFORMATION

The EU Action Plan Against Disinformation, presented by the European Commission in December 2018, consisted of a set of initiatives to reinforce capacities and strengthen cooperation between Member States and EU institutions in order to proactively address disinformation.[297] The main goal was to develop a coordinated response to the related challenges, especially in view of the 2019 European Parliament elections.[298]

As a key pillar of this Plan, a self-regulatory Code of Practice was agreed upon by the European Commission, major social media companies and advertisers.[299] In October 2018 it was signed by the online platforms Facebook, Google and Twitter, Mozilla, as well as by advertisers and other advertising industry actors, who presented their roadmaps for its implementation. In May 2019, Microsoft also subscribed to the Code of Practice and presented its own roadmap (the Code was later updated in 2022).

The Action Plan against Disinformation focused on four areas:

- **Improved detection:** reinforcing EU institutions and delegations in neighbourhood countries with significant additional specialized staff and data analysis tools.
- **Coordinated response:** putting in place a dedicated Rapid Alert System, involving the EU institutions and Member States, for the sharing of data and assessments of disinformation campaigns, as well as to issue real-time alerts on disinformation threats.
- **Online platforms and industry:** implementation of commitments agreed upon under the Code of Practice, such as ensuring transparency of political advertising, strengthening actions to close fake accounts, labelling messages disseminated by bots and collaborating with fact-checkers and academia to detect disinformation campaigns and increase the visibility and reach of verified content. The European Commission, in cooperation with the European Regulators Group for Audiovisual Media Services (ERGA), closely monitored the implementation of the commitments.
- **Raising awareness and empowering citizens:** The EU institutions and Member States promoted media literacy both through awareness-raising campaigns and dedicated programmes. National multidisciplinary teams of independent fact-checkers and researchers were supported, to identify and uncover online disinformation campaigns.

---

297 See: https://eeas.europa.eu/headquarters/headquarters-homepage/54866/action-plan-against-disinformation_en and https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6647.

298 The Action Plan and the related Code of Practice were part of a series of initiatives implemented to follow-up on the adoption of the European Parliament's Resolution on EU strategic communication to counteract propaganda against it by third parties (2016/2030(INI)). These efforts also included the set-up of the European Commission's High Level Group on 'Fake News' and Online Disinformation (March 2018) and the EC Communication on 'Tackling online disinformation: a European Approach' (April 2018).

299 See: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6647.

FIGURE 9: THE EU ACTION PLAN AGAINST DISINFORMATION[300]

To ensure continuous monitoring of the implementation of the commitments under the Code, Facebook, Google, and Twitter agreed to report on their actions on a monthly basis – including in relation to enhancing ad placement scrutiny, transparency of political and issue-based advertising, and tackling fake accounts and the malicious use of bots.

When assessing the Code's implementation after one year, the European Commission found that the 2019 European Parliament elections were not free from disinformation. Actions taken by the EU – together with numerous journalists, fact-checkers, social media platforms, national authorities, researchers, and civil society – contributed to narrowing down the space for foreign interference and coordinated campaigns to manipulate public opinion. Still, they did not manage to eliminate all of it.

The Commission also noted, among the positive outcomes of the experience, the improved transparency and closer dialogue with platforms regarding their policies against disinformation.[301]

However, the Code did not resolve independent researchers' access to the data they need. The initiative has shown shortcomings, including the limited number of the Code's signatories, insufficient provision of data access, and lack of consistency in the data sets retrieved – which hampers the possibilities of carrying out coherent research. Furthermore, the scope of the actions undertaken by each Internet intermediary to implement their commitments also varied significantly.

The experience revealed the limitations inherent to self-regulation, which has prompted recommendations for the establishment of a related body that would oversee and enforce rules to be implemented by social media networks, particularly during elections. Another recommendation was for the multi-stakeholder approach to design and implement these rules to be substantially broadened. While the Code generally represents a step in the right direction, much more needs to be done.[302]

---

[300] See: https://ec.europa.eu/commission/sites/beta-political/files/factsheet_disinfo_elex_140619_final.pdf.

[301] See: https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2020)180&lang=en.

[302] A. Kuczerawy, "Fighting Online Disinformation: Did the EU Code of Practice Forget about Freedom of Expression?", Forthcoming in: "Disinformation and Digital Media as a Challenge for Democracy" European Integration and Democracy Series, 6, 2019. Available at: https://ssrn.com/abstract=3453732, as cited by Posetti.

5.3.2. EU CODE OF CONDUCT ON COUNTERING ILLEGAL HATE SPEECH ONLINE

A previous experience, also at the EU level, was the Code of Conduct on Countering Illegal Hate Speech Online,[303] to which Facebook, Microsoft and Twitter agreed in May 2016. In addition, Instagram, Snapchat, and Dailymotion joined this initiative in January 2019, and TikTok announced its participation in September 2020.[304] Through this self-regulatory Code, the companies assigned themselves, amongst others, the responsibility to remove "illegal hate speech", as well as to put in place community standards and review valid notifications of online hate speech within 24 hours. However, they agreed to review requests for removal of content according to "their rules and community guidelines and, where necessary, national laws", which oftentimes may not coincide with the standard expressed in the ICCPR, or the ECHR. For instance, Facebook's Community Standards refer to "objectionable content"[305] or content that is "cruel and insensitive",[306] which are not the benchmarks laid down in international law. The concept of incitement to hatred or discrimination prohibited by Article 20 of the ICCPR, also seems to be missing in the vocabulary and concepts included in Facebook's Community Standards.

The Code's implementation is assessed regularly by the European Commission in collaboration with a network of organizations located in different EU countries. An information note produced by the EC on the advances made in combating hate speech through this initiative from 2016 to 2019 finds that:

> "[it] has contributed to achieve quick progress, including in particular on the swift review and removal of hate speech content (28% of content removed in 2016 vs. 72% in 2019; 40% of notices reviewed within 24h in 2016, 89% in 2019). It has increased trust and cooperation between IT Companies, civil society organisations and Member States authorities in the form of a structured process of mutual learning and exchange of knowledge. This work is complementary to the effective enforcement of existing legislation (Council Framework Decision 2008/913/JHA) prohibiting racist and xenophobic hate crime and hate speech and the efforts needed by competent national authorities to investigate and prosecute hate motivated offences, both offline and online."[307]

However, similar critiques to those made in relation to the EU Code of Practice on Disinformation have been directed at the Code of Conduct on Countering Illegal Hate Speech Online, including the concerns about the risks to freedom of expression when placing companies in a position of enforcing measures that can limit speech online,[308] particularly as Internet intermediaries may "sanitize speech" in the interests of avoiding lawsuits, bad advertising, and loss of customers.

5.3.3. FACEBOOK'S OVERSIGHT BOARD

Social platforms, including Facebook, Twitter, TikTok, Instagram, Snapchat, and others, have put in place complaints and appeals mechanisms, which function based on the agreement individuals sign to use their services. In 2020, Facebook– going beyond its existing moderation appeals

---

[303] Code of Conduct on Countering Illegal Hate Speech Online (2016), adopted by Twitter, Facebook, Microsoft and YouTube. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en#theeucodeofconduct.

[304] Ibid.

[305] See: https://www.facebook.com/communitystandards/objectionable_content.

[306] Ibid.

[307] See: https://ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play__0.pdf.

[308] Coche lists the terms which refer to this practice as: "non-law based voluntary enforcement measures", or "intermediarization". E Coche, Privatised enforcement and the right to freedom of expression in a world confronted with terrorism propaganda online. ***Internet Policy Review***, 7(4), 2018.

process – created an Oversight Board composed of highly renowned human rights and freedom of expression specialists.[309] The main idea is to have an independent body to handle the most complex moderation decisions, initially in the narrow remit of what content is taken down by the company. Its 20 first members were announced in May 2020,[310] and the board is expected to continue expanding to include up to 40 members.

The Oversight Board oversees the appeals from users who have seen their content being removed from Facebook platforms and on cases referred to it by Facebook. For claims submitted either by a user or by Facebook itself, the Board has 90 days to complete the adjudication of the process, including the publication of the decision and a policy recommendation for Facebook. All users have the option of appealing to the Board as an extension of the Facebook moderation appeal process that is already in place. The Board publishes all its decisions, and Facebook is committed to issuing a public response regarding them.[311]

Concerns have been raised about the Oversight Board, noting, for example, that it will not have any control over the algorithms that Facebook uses to manage content. Critics also point out that the Board's powers are restricted. It cannot request internal documents from Facebook, and it only reviews a limited number of selected cases, aiming to set precedents that are likely to affect many users.[312] A group of critics has launched a "Real Facebook Oversight Board",[313] a mechanism to act as a watchdog in relation to the company's content moderation policy.[314]

5.4. THE RELEVANCE OF VOTER EDUCATION AND MEDIA INFORMATION LITERACY

The development of media and information literacy (MIL), as elaborated further in 5.4.2 below, should accompany any regulatory, self-regulatory or co-regulatory approach to online content. The proliferation of dis- and misinformation depends on people's inability to distinguish between true and false, which is why a key part of the solution lies in critical thinking. Especially during electoral periods, citizens need to be able to assess whether certain news content is reliable or not, and to distinguish fact from opinion. This would underpin more informed choices about the news they consume and the content they share, comment on, or reutilize. The strengthening of critical skills should go beyond textual content, and also extend to building the electorate's awareness about manipulation strategies that appeal to their emotions and their digital rights/related rights online as well as offline.[315]

A second edition of UNESCO's model curriculum for educators and learners on Media and Information Literacy, titled "Media and Information Literate Citizens - Think Critically, Click Wisely"[316], published in 2021, highlights the value of literacies not only about media and information, but also about privacy and the political economy of Internet communications

---

[309] See: https://www.theguardian.com/technology/2020/may/07/will-facebooks-new-oversight-board-be-a-radical-shift-or-a-reputational-shield.

[310] See: https://about.fb.com/news/2020/05/welcoming-the-oversight-board/.

[311] The most prominent case considered by the Board so far was the ban on former President Donald J. Trump which resulted in the upheld of the Facebook's decision, thus preventing any immediate return by Mr. Trump to mainstream social media and renewing a debate about tech power over online speech.

[312] See: https://time.com/5918499/facebook-oversight-board-cases/.

[313] See: https://the-citizens.com/about-us/.

[314] See: https://www.reuters.com/article/us-facebook-oversight/facebook-critics-launch-rival-oversight-board-idUSKCN26G1R6?il=0; https://www.theverge.com/2020/9/25/21454094/facebook-oversight-board-election-criticism-activists.

[315] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***.

[316] See: https://en.unesco.org/news/media-and-information-literate-citizens-think-critically-click-wisely.

companies, which are both highly relevant to elections.[317] Media and information literacy can also help individuals to identify hate speech and learn how they can contribute to counteracting it online.[318]

Thus, going beyond regulation of online content and political advertising, governments could consider modifying or integrating a media and information literacy component into EMBs' voter or citizen education mandates, besides including media and information literacy in school curricula. EMBs should cooperate with CSOs and educational institutions in this area and can also partner with media to develop and disseminate counter-narratives to disinformation during electoral processes, as well as to prevent and combat hate speech.

In most democracies, EMBs are responsible for ensuring that citizens receive this basic information and that it is presented in a non-partisan manner. In some systems, administrators may be assisted by civil society groups or CSOs. Voter education is key to the integrity of the electoral process and covers all phases of election process, including voter and candidate registration processes.

In the current context, voter education and media and information literacy should be closely interrelated. It is very relevant for voters to understand the dangers of digitally spread disinformation and hate speech and have basic fact-checking tools. EMBs should therefore consider integrating media and information literacy topics into their voter education programs in collaboration with other state agencies, CSOs, different types of media, educational institutions, social media platforms, and international organizations.

5.4.1. VOTER EDUCATION

In many democracies voter education (VE) starts in elementary school as a feature of basic civic education programs. This prepares students to understand the part they can fulfil in a democracy when they become eligible to vote.[319]

Voters need to be familiar with the constitution and electoral legal framework of their country so that they can meet their obligations in a responsible manner. Informed, responsible voters help safeguard electoral integrity. They do not make false statements that might disrupt or prevent an election. They do not act illegally, intimidate other voters or try to tamper with the election results. They turn out to vote in an election because they understand the importance of participating to the democratic system. Without appropriate education, it can be hard to prevent vote buying or vote tampering through intimidating tactics, especially in countries with high unemployment, low incomes and security problems. Voters may not be aware of their rights or the mechanisms that are used to protect the secrecy of their vote, or of what motivates vote buyers.

Voter education should provide factual information so that voters can participate knowledgeably. In a neutral way, programs should explain when, where and how to register and vote; the documents that must be shown; and how to mark a ballot so that it is valid and will be counted. If voter education has political content, it becomes propaganda that may sway opinion and is intended to build support for a specific candidate or platform.

5.4.2. MEDIA AND INFORMATION LITERACY (MIL)

Media and information literacy is an umbrella term that encompasses various competencies that enable individuals and groups to navigate the turbulent seas of today's information and communications environment. It covers a large spectrum of knowledge, skills, attitudes, and values.[320]

---

[317] See: https://en.unesco.org/themes/media-and-information-literacy.

[318] UNESCO, 2015, ***Countering online hate speech***. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000233231.

[319] See: https://aceproject.org/ace-es/topics/ei_new/eif/eif05.

[320] UNESCO, Think Critically, ***Click Wisely! Media and Information Literate Citizens***, Second edition of the UNESCO

Media and information literacy enhances the capacity of citizens to critically and meaningfully engage with information, including in achieving the Sustainable Development Goals, while claiming and enjoying their fundamental human rights, such as those expressed in Article 19 of the Universal Declaration of Human Rights, which includes the freedom to hold opinions without interference, and to seek, receive and impart information and ideas through any media and regardless of frontiers.

The main benefits of media and information literacy are:

1. Media and information literacy offers a sustainable way to tackle the rise of disinformation.
2. Through media and information literacy, people can self-empower to understand how they can interact critically with and use media and digital tools to create positive outcomes, and thereby ensuring a better Internet and contributing to upholding the vision of information as a public good.
3. Media and information literacy is a prerequisite for other forms of literacy, such as health literacy, science literacy, cultural literacy, global citizenship education and education for sustainable development.
4. When media and information literacy is integrated in all types of formal, informal and non-formal learning, including voter education, it helps to defend against privacy infringements and enables all people to respect the privacy rights of others.
5. Media and information literacy becomes a necessary skillset when involved in AI ethics and the ever-evolving digital transformation processes.
6. Media and information literate persons are more likely to reject unvalidated information, biases and stereotypes that create or reinforce existing inequalities for example gender inequalities, or discrimination based on age, race, socioeconomic status, religion, ethnicity, nationality etc.
7. Media and information literacy offers a way to build citizens' and users' capacity to assess what merits trust, at a time when this is eroding.
8. A society that is media and information literate fosters sustainable development and the development of free, independent and pluralistic media, as well as open information and digital communications systems.

5.4.3. YOUTH PARTICIPATION

Despite constituting one-fifth of the world's population[321] and often being engaged in activism and informal processes that are politically relevant, young people aged between 15 and 25 are not significantly represented in parliaments and other political institutions. A large proportion among them do not take part in elections either.[322] This can impact the quality of democratic governance. Youth should, therefore, be a key audience of outreach actions, voter education and media and information literacy initiatives ahead of elections. These efforts are key to youth empowerment, in line with the UN Youth 2030 Strategy, which, among other areas of action, fosters youth's participation in political and public affairs, by leveraging "the capacity of the UN to promote young people's right to participate in public affairs, including in political and civic processes, platforms and institutions at all levels, such as elections, constitution-making processes, political parties and parliaments.[323]

Media and Information Literacy curriculum for educators and learners, 2021. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000377068.

[321] UNDP, ***Youth, Political Participation and Decision-Making***. Available at: https://www.un.org/esa/socdev/documents/youth/fact-sheets/youth-political-participation.pdf.

[322] UNDP, ***Enhancing Youth Political Participation throughout the Electoral Cycle***, October 2015, p. 3, https://www.undp.org/content/undp/en/home/librarypage/democratic-governance/electoral_systemsandprocesses/enhancing-youth-political-participation-throughout-the-electoral.html.

[323] UN Youth 2030 Strategy, p. 12,

In this regard, websites have been a useful platform for EMBs to interact with young people in different countries and have been especially relevant in contexts where those voting for the first time must register to do so. The Electoral Commission in the United Kingdom, for example, set up a website for educational purposes, with young people as its main target audience and including resources specially designed for this sector of the population.[324] Media and Information Literacy efforts targeting young people are also critical, as they are the most active users of social media. For example, in Finland,[325] recently rated Europe's most resistant nation to disinformation, primary school students are taught how to identify disinformation and combat it.[326] Moreover, the strategic use of innovative approaches, combining dissemination of voters' education via social media with more traditional, face-to-face methods, are also important when reaching out to youth.

For instance, during the 2018 and 2020 USA elections, social media platforms reached a very large segment of young people, many of whom were potential first-time voters with positive effect on youth voter turnout, especially for those youth who lacked election information and outreach from other sources.[327]

While social media companies have supported voter registration and engagement in the past, during the 2018 and 2020 US elections many of them expanded their efforts to provide accurate information about voting and the election in an accessible way, and to attempt to encourage young people to vote.

For instance, Instagram ran a campaign to encourage users to register to vote before the midterm elections in 2018. Snapchat ran a similar campaign and reminded users to vote on election day—along with providing a map to help users get to their polling place. These digital initiatives were a valuable contribution to the collective work of voter education that other groups do both online and in-person, because they can provide information easily and at scale even when a potential voter is not actively looking for election-related information or is not connected to an organization that would provide it.

Some of the beneficial effects of social media for youth during elections are [328]:

- **Social media platforms have extraordinary reach:** social media can potentially integrate voting and election information into people's social lives, thereby normalizing electoral participation and promoting a culture of political engagement. Voting, then, becomes social—an experience young people can use to encourage others to do the same.
- **Social media platforms are reaching youth not engaged by candidates and campaigns:** social media platforms that share registration and voting information may be serving as a crucial complement, reaching youth that other efforts do not.
- **Social media may be a particularly relevant source of information for first-time voters:** social media may play an important role for young people living in rural or non-urban areas, where traditional outreach groups may not reach them as easily.
- **Youth who get election information are more likely to vote:** the best ways to promote youth voting is to reach out to young people and give them information and make them feel involved in decision-making.

https://www.un.org/youthenvoy/wp-content/uploads/2018/09/18-00080_UN-Youth-Strategy_Web.pdf.

[324] See: https://www.electoralcommission.org.uk/welcome-your-vote/resources-14-18-year-olds.

[325] CNN Special Report, ***Finland is winning the war on 'fake' news. What it's learned may be crucial to Western democracy***, https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/.

[326] The Guardian, 29 January 2020, ***How Finland starts its fight against 'fake' news in primary schools***, https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news.

[327] See: https://circle.tufts.edu/latest-research/five-takeaways-social-media-and-youth-vote-2018.

[328] See: https://circle.tufts.edu/latest-research/five-takeaways-social-media-and-youth-vote-2018.

- **Social media can help to strengthen civic and voter education and encourage political participation among youth:** it's far more likely that young people who already have an interest in elections and/or political issues seek election information on social media and use these platforms to deepen their engagement by connecting with peers, organizations, and candidates. Therefore, social media can potentially help move young people from "intent" to "action".[329]

BOX 14: *YVOTE* KENYA

YVote was a joint effort launched by Kenya's Electoral Assistance Program (KEAP) and the International Foundation for Electoral Systems (IFES) in preparation of the 2017 general elections.[330] The initiative was set out to reach the youth of Kenya to trigger higher voter registration and voter turnout in the general elections. The initiative originated from a lack of voter registration from a particular share of the population – the youth.

The main target audience of YVote – the young, poor, and politically marginalized population – consists of youth between the age of 18-29 years with a Living Standard Measurement (LSM) of 2-8. The LSM is a concept that was created by the South African Advertising Research Foundation and categorizes individuals within the population according to their ability to access certain products such as electricity, fridges, or microwaves.[331]

Population categorized with an LSM of around 2-8 are rather poor. This group had shown lack of participation in Kenya's elections before 2017. This group makes up nearly 70% (also including youth under 18) of the population and hence accounts for most eligible voters. Therefore, YVote was created to target that audience and to create awareness about the elections, their ability to influence with their votes and how the election procedures work as an election integrity building measure. YVote utilized a digital strategy to engage with the youth. For that, YVote purchased and placed social media ads, and provided videos on TV that educated about election processes such as counting, tallying and transmission of results. Special training and events were held for underserved or marginalized groups in Kenya's society, including women and persons with disabilities.

BOX 15: YOUTH-LED MONITORING OF ELECTIONS USING AGGIE

Aggie is an open-source social and online media aggregation tool[332] that enables the monitoring of a high volume of social media traffic from several sources. The tool was first developed to support youth-led, grassroots social media monitoring of the 2011 Nigerian general elections. It includes trend visualization and search features that, according to the team behind it, have helped detect unfolding events more quickly than the traditional media and authorities. Aggie has also been used to monitor elections in Ghana and Argentina.

The research team behind Aggie includes civic technologists from Georgia Tech and

---

[329] ***Expanding the Electorate. How Simple Changes in Election Administration Can Improve Voter Participation Among Low-Income Youth***, Center for Information and Research on Civic Learning & Engagement, 2018.

[330] R. Nackerdien, 2017, ***YVote - Youth Outreach Campaign for the 2017 General Elections***, IFES.

[331] P. Haupt, 2017, The SAARF Universal Living Standards Measure (su-lsm™) - 12 Years of Continuous Development. Retrieved from http://saarf.co.za/LSM/lsms.asp#.

[332] See: https://www.getaggie.org.

Sassafras Tech Collective (based in the USA), who were contacted by the Nigerian youth group "EnoughIsEnough" for their support in developing software to monitor content related to the election on Twitter. This collaboration gave way to Aggie, the social media monitoring tool, as well as to the set-up of a Social Media Tracking Center, which is the physical space from where a group of volunteers carry out continuous social media monitoring. There are usually three teams involved in the implementation of this methodology: a relevance team (which observes trends, scans reports and "creates incidents" when content is judged worthy of further attention), a veracity team (which verifies incidents and passes those that are confirmed to the escalation team), and an escalation team (which liaises with authorities so that proper action is taken).

The creators of Aggie refer to their efforts as "social election monitoring", which differentiates them from more formal ways of monitoring carried out by trained teams of observers. This type of tool allows for monitoring by domestic groups with fewer resources.[333] However, Aggie teams have also collaborated with formal monitoring efforts to triangulate work, which they found led to a higher percentage of resolved or closed. A new initiative with similar facilities is UNDP's iVerify tool.

5.5. BUILDING CAPACITIES AMONG JUDICIAL ACTORS

The judiciary is pivotal to the integrity of the electoral processes[334] and an important player all along the electoral cycle, although it can be said that the judiciary is reactive rather than proactive, meaning that judges can only decide on what is placed before them. The judiciary has a role in overseeing the rules of the game (legislation) by ensuring the compatibility of the electoral law with the Constitution, that a level playing field is provided to all parties, and that any violations of these rules are addressed, and the appropriate redress is afforded to concerned parties. Here, the judiciary plays the role of enforcer or referee ensuring that the elections are genuinely conducted in line with international standards. This becomes particularly important when the outcomes of opaque elections lead to violence and are often tainted by allegations of fraud, thus perceived as illegitimate by the losing parties or the public. In such cases, an impartial judiciary can hear and peacefully resolve electoral disputes, ensure accountability, and help uphold the rule of law. Judicial independence and separation of powers can protect against undue influence from the other branches of the government and lend credibility to its decisions.[335]

For example, in 2016, the Austrian Constitutional Court nullified the presidential elections as it found that the confirmed irregularities had affected a number of votes that would in theory be enough to change the election result.[336] In turn, the Supreme Court of Kenya nullified the results of the presidential elections in 2017 because it found that the polls were "neither transparent nor verifiable", and that they were not in line with the Constitution as the election result was declared before all results from polling stations had been received.[337] In May 2020, the Malawi Supreme Court upheld an earlier court ruling that annulled the presidential election that had been held in

---

[333] See: https://www.getaggie.org/papers/ictd2016.pdf.

[334] The right to due process and a free and fair trial by an impartial tribunal is a right that pertains to elections even though it does not directly derive from Article 25 of the ICCPR.

[335] S. Tamboly, ***The Role of the Judiciary in Preventing Post-Electoral Violence***, International Development Law Organization (The Hague), http://www.idlo.int/, https://ihrp.law.utoronto.ca/role-judiciary-preventing-post-electoral-violence.

[336] See: https://www.theguardian.com/world/2016/jul/01/austrian-presidential-election-result-overturned-and-must-be-held-again-hofer-van-der-bellen.

[337] See: https://www.theguardian.com/world/2017/sep/20/kenyan-election-rerun-not-transparent-supreme-court.

May 2019, due to significant irregularities,[338] and a new election was held in June 2020.

The role of the judiciary is also crucial in upholding freedom of expression online, especially during the campaign period. The relevance of this role was referred to repeatedly in reports by the UN Special Rapporteur on Freedom of Opinion and Expression, which highlighted that "an independent body should be in charge of enforcing any legislation restricting the right to freedom of expression, in a non-discriminatory or arbitrary manner, and with sufficient safeguards against abuses, including the possibility of challenge and remedy against abusive application".[339] Further, "States should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy."[340] The courts should therefore be the final port of call on issues such as whether freedom of speech has been respected or whether an act of hate speech has taken place during the election campaign. Such ultimate assessment should not be made by a social media platform or another Internet Service Provider. In the context of the Council of Europe, for example, the courts have ruled in several "notify and take-down" cases concerning hate speech, to guide the action of social media platforms and safeguard the right to freedom of expression.[341]

Given the fast development of the Internet, social media, social messaging and AI, their implications for human rights and the complexities they entail, it is important to build awareness and capacities among judiciary actors on these matters.

BOX 16: UNESCO'S JUDGES INITIATIVE

First launched in 2013, UNESCO's Judges Initiative seeks to strengthen judicial actors' awareness and knowledge of international standards and regional jurisprudence on freedom of expression, access to information and digital challenges related to the Internet and the safety of journalists. It has taken advantage of the expanded opportunities offered by ICTs, being implemented mainly through a series of online courses, which were accompanied by workshops and seminars.

Since the start of this effort, more than 23,000 judicial actors have been trained in over 150 countries in the world. To implement these efforts, UNESCO has signed cooperation agreements with the African Court on Human and Peoples' Rights, the Economic Community of West African States Court of Justice, and the Inter-American Court of Human Rights, as well as with associations of Chief Justices and Attorney Generals.[342]

Under the umbrella of this initiative, UNESCO has for example developed an online course on AI and the Rule of Law, in cooperation with the Institute of Electrical and Electronics Engineers (IEEE) and UNESCO's Category 2 Centre, CETIC.br – at the Brazilian Network Information Center (NIC.br) – among other partners.[343] This online course will reinforce judicial operators' capacities to tackle AI-related issues in their domain of work, equipping them with information and knowledge concerning for example the use of AI

[338] See: https://www.aljazeera.com/news/2020/05/malawi-court-rejects-president-appeal-poll-annulment-200508140237123.html.

[339] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2011, para. 69, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G11/132/01/PDF/G1113201.pdf?OpenElement.

[340] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2018, para. 66, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement.

[341] ***Delfi AS v. Estonia*** (Application No.64569/09), Judgement of the European Court of Human Rights of 16 June 2015; ***MTE v. Hungary*** (Application no. 22947/13), Judgement of the European Court of Human Rights of 2 February, 2016, Final 02/05/2016.

[342] See: https://en.unesco.org/training-foe.

[343] See: https://en.unesco.org/news/unesco-launches-survey-judicial-operators-ai-and-rule-law.

systems for investigative assistance and automating decision-making processes, as well as international human rights law as it concerns AI and related implications and risks, including in relation to free and fair elections.[344]

BOX 17: MEXICO'S ELECTORAL TRIBUNAL OF THE FEDERAL JUDICIAL BRANCH

In 2014, in the framework of a broader set of electoral reforms in the country, the Electoral Tribunal of the Federal Judicial Branch (TEPJF) was given jurisdiction over complaints pertaining to electoral campaigning and the media.

As Mexico did not have legislation regulating cybersecurity and elections,[345] this meant that, even in the absence of rules specifically focused on social media use, disinformation campaigns and other digital challenges, the TEPJF had to adopt binding decisions. It did so during the electoral process in 2018. It issued important rulings aimed, for instance, at safeguarding freedom of expression while also supporting the work of fact-checkers to counter disinformation.

BOX 18: DIGITAL PLATFORMS ADS MANAGEMENT

In response to criticism on how it has handled paid political online ads (allowing, for example, dark ads), Facebook established an Ad Library in 2018 in the form of a searchable website that anyone can access. It keeps a record of every active and inactive ad about social issues, elections, or politics that has been run since March 2019, and the intention is to keep all the related threads for seven years. The library aims to provide advertising transparency "by offering a comprehensive, searchable collection of all ads currently running across Facebook products, including Instagram."[346]

Facebook requires those wishing to advertise on its platforms to provide verifiable public contact details before they can run political campaigns on them. The restrictions mandate advertisers on political topics (defined differently in each nation) to prove that they live in the country that they are targeting, and to store all their adverts in a public database, along with information about targeting, expenditure and reach.

Advertisers should disclose who paid for each advert – a requirement for which Facebook was criticized, since it allowed users to enter whatever information they wanted in the box and did not verify the names that were provided. Following such criticism, a change was introduced, so that although users are still allowed to write what they want when it comes to the source funding, they now must provide at least a phone number or email address through which interested parties can contact them. Users who advertise in a personal capacity are free not to enter that contact information, but their name is published instead, as verified by the site.

Regarding Google's ad policies, all ads must be "clear and honest, and provide the

[344] See: https://en.unesco.org/artificial-intelligence/mooc-judges.

[345] See: https://www.ifes.org/news/electoral-justice-assessment-2018-mexican-elections.

[346] For more information on the Ad Library, see https://www.facebook.com/help/259468828226154.

information that users need to make informed decisions."[347] Google prohibits ads that deceive users by including misleading information about products, services, or businesses, including "deceptively doctoring media related to politics, social issues, or matters of public concern."[348] From September 2020, the company's misrepresentation policy for ads also "prohibits accounts from coordinating with other sites or accounts to conceal or misrepresent their identities or other material details if the content of the account relates to politics, social issues, or matters of public concern."[349]

BOX 19: POLITICAL ADVERTISEMENT AND MICROTARGETING IN THE USA

In November 2019 Twitter decided to ban all political advertising and Google also announced a ban on adverts targeting people based on their political party, however it still allows gender, age and location-targeted ads. In turn, Meta, which also owns the popular social media platform Instagram and the messaging application WhatsApp, announced that it would reconsider its micro-targeting policy to increase the minimum target size of political advertisements yet it decided not to fact-check ads from political campaigns, maintaining that it should not be the arbiter of truth. The company faced criticism for this approach as many saw it as being insufficient for countering disinformation in electoral campaigning contexts. Critics had also voiced their concerns regarding a different decision made by Facebook earlier on, during the 2019 UK election campaign, when the company deleted a Conservative Party election ad that used BBC News footage arguing that it infringed the BBC's intellectual property (IP) rights, which resulted in accusations of mingling in the electoral campaign.[350]

In the recent years, major social media companies modified their policies and practices. Nevertheless, regarding electoral-related disinformation narratives that questioned voting procedures or election integrity, a fundamental lack of transparency raised concerns about whether these policies and practices were effectively implemented. For instance, this lack of transparency is related to some companies' changes in their algorithmic ranking and recommendation systems to prevent election-related misinformation and disinformation amplifying. While this was a positive move, it was impossible to independently verify what impact these disclosed changes had on the spread of misleading and false election-related information and how or if humans were involved in the policy implementation processes.[351]

---

[347] See: https://support.google.com/adspolicy/answer/6008942?hl=en.

[348] See: https://support.google.com/adspolicy/answer/6020955.

[349] See: https://support.google.com/adspolicy/answer/9991401?hl=en#:~:text=In%20September%202020%2C%20the%20Google,or%20matters%20of%20public%20concern.

[350] C. Newton, 2020, ***Facebook's revised political advertising policy doubles down on division***, https://www.theverge.com/interface/2020/1/10/21058616/facebook-political-ads-targeting-misinformation-polarization; I. Togoh, 2019, ***Facebook hints at plans to restrict controversial microtargeted political ads – report***. Retrieved from: https://www.forbes.com/sites/isabeltogoh/2019/11/22/facebook-hints-at-plans-to-restrict-controversial-micro-targeted-political-adsreport/#4db987ac6714; BBC online, 2019, ***General election 2019: Facebook bans Tory ad over BBC footage***, https://www.bbc.com/news/election-2019-50624086.

[351] See: https://www.newamerica.org/oti/reports/protecting-vote/executive-summary.

5.6. ADDRESSING VIOLENCE AGAINST WOMEN IN ELECTIONS

The Internet, social media, AI, and other technologies can be used to tackle the challenges related to gender equality and the dangers faced by women during election periods for example through the implementation of social media monitoring, or by facilitating the use of social media for women to report physical violence as well as threats or harassment made online or through private messaging services.

In the publication Violence Against Women in Elections (VAWIE) Framework, it is noted, for example, that:

> "low-cost or public-domain software services such as Hootsuite, TweetReach, Klout, and Social Mention that are now being used to monitor social media traffic may be used to track ICT-based violence directed at women political activists, candidates and politicians. Open source software mash-ups such as Ushahidi can map data collected from SMS, Twitter, Facebook, YouTube, phone calls, and email. Reports of events can be seen on the website map in near real-time, depending on the resources for data processing."[352]

These tools permit instances of violence to be mapped in close to real-time, and the fact that the process is anonymous may increase the number of reports made. Technological tools can also be used in support of awareness-raising and advocacy against gender-based violence. For instance, in Argentina, UN Women and Instagram developed a Safety Guide for Women in Politics aimed at protecting candidates and creating awareness[353].

Tackling violence against women in elections calls for cooperation among EMBs, security forces, media actors, electoral observers and organizations involved in combating gender-based violence, the judiciary, political parties and candidates, among others. Sensitizing and training these actors on gender-based violence during elections is thus critical.[354]

In 2021 UN Women published a Guidance Note providing technical advice to UN Women and country teams on how they can support Member States to address violence against women in politics (VAWP)[355]. It draws on existing definitions, insights and framing generated from research, normative advancements and programmatic collaboration, including: the Report of the Secretary-General for the 65 Commission on the Status of Women on Women's full and effective participation and decision-making in public life, as well as the elimination of violence, for achieving gender equality and the empowerment of all women and girls (E/CN.6/2021/3); key messages for the UN system on VAWP adopted by the UN Executive Committee in 2020 (Annex A); a thematic report of the UN Special Rapporteur on violence against women, its causes and consequences on VAWP submitted to the 73rd regular session of the UN General Assembly (2018), UN Women and UNDP's programming guide on Preventing violence against women in elections; two expert group meetings and a global mapping of lessons learned and good practices of UN Women Country Offices (COs).

The UN Women 2021 Guidance Note therefore focuses primarily on women in politics but it likewise applies to violence against women in public life more broadly, including that perpetrated against women human rights defenders, journalists, those active in civil society and in other areas of public life. It can also be used to guide and inform the work of other United Nations agencies and development partners.

---

[352] IFES VAWIE: https://www.ifes.org/sites/default/files/vawie_framework.pdf, p. 10.

[353] See: https://www.clarin.com/entremujeres/genero/-feminazi-hueca-frigida-instagram-lanzo-guia-genero-mujeres-politicas_0_YG8TvI27f.html.

[354] See: https://www.ifes.org/sites/default/files/vawie_framework.pdf.

[355] For more details see ***Preventing Violence against Women in Politics: Guidance Note***, UN Women, 2021. Available at: https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2021/Guidance-note-Preventing-violence-against-women-in-politics-en.pdf.

BOX 20: IFES' VIOLENCE AGAINST WOMEN IN ELECTIONS (VAWIE) SOCIAL MEDIA ANALYSIS TOOL

This tool, developed by The International Foundation for Electoral Systems (IFES), uses AI-based data analysis to measure the gendered dimensions of online electoral violence against women and to understand the factors that drive it, by helping to detect trends and patterns. It does so via data scraping, data mining and analysis, looking at information available on online public spaces. Data is extracted from a large volume of Facebook and Twitter posts (including for instance in public social media accounts of candidates), webpages and public forums; it is filtered by using particular words, subjects, dates, user, popularity, etc., to find patterns; and it is interpreted and classified.[356]

The analysis that the tool facilitates can contribute to answering questions related to whether women are impacted disproportionately by online violence and harassment, if the types of violence change over time, their drivers and causes, their targets, their perpetrators, the digital platforms through which these instances of online violence flow and the speed of their dissemination, among other relevant aspects. The tool was first piloted in Sri Lanka, Ukraine and Zimbabwe in 2018 and 2019.

As noted by IFES, EMBs hold a privileged position to detect incidents and respond to them with this tool, as they possess the list of candidates in an election and can work with them to gather further information sources. They can also allow women candidates to share information about the cases of online and physical violence they experience and join forces with the executive and legislative powers toward the adoption and effective enforcement of laws protecting women from online violence during elections. The Paper presenting the tool also outlines the roles of legislators, law enforcement and security forces, political parties, media, social media companies and Internet governing bodies, advocates, and civil educators.[357]

BOX 21: THE #THINK10 PLANNING TOOL

The #think10 safety planning tool provides women in politics guidance on how to enhance their personal security by combining scores from a self-assessment questionnaire and the country score from National Democratic Institute's new Women's Political Participation Risk Index (WPPRI). The WPPRI calculates the risk for politically active women in 172 countries. In using the tool, women in politics can develop a safety plan relevant to their personal and professional profile, and in their political context. Each country's ranking in the WPPRI is built on three indicators: the level of women's political participation at the national level; the state of democracy in each country; and the likelihood of violence that women in that country face. NDI has based these indicators on data gained from the Inter-Parliamentary Union, the Economist Intelligence Unit, and Georgetown University's Institute of Women Peace and Security.[358]

---

[356] For more details see ***Violence Against Women in Elections Online: A Social Media Analysis Tool***, IFES, 2019, pp.12-14. Available at: https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf.

[357] Ibid. For details, see pp. 51-53.

[358] See: https://www.ndi.org/publications/ndi-launches-think10-groundbreaking-safety-planning-tool-designed-safeguard-women.

ACTIVITY V

The following activity has the objective of determining the participant's level of knowledge on which measures EMBs can take in order to tackle disinformation all along the electoral cycle, in the short, medium and long term; media regulation during elections; regulation, self-regulation, and co-regulation of online content; codes of practice agreed upon by Internet intermediaries; the relevance of voter education and media information literacy; building capacities among judicial actors and addressing violence against women in elections.

Suggested guiding questions for a discussion:

I. Can you identify the four main categories to tackle disinformation?
II. Which preventive measures can be taken? Please explain.
III. Why identification and monitoring measures are extremely relevant during electoral periods? Can you provide any examples?
IV. What are the differences between regulation, self-regulation, and co-regulation of online content?
V. Codes of practice agreed upon by Internet intermediaries are also relevant for democracy. Please, explain.
VI. What does EU Action Plan Against Disinformation, presented by the European Commission in December 2018, consist of?
VII. What is Facebook's Oversight Board?
VIII. Please explain the relevance of voter education and media information literacy particularly for the youth.
IX. Why building capacities among judicial actors is critical to ensuring freedom of expression during elections?
X. How to effectively address online violence against women during elections?

6. GOOD PRACTICES AND GUIDANCE

OBJECTIVES OF THIS SECTION

- Provide an overview of tools for electoral administration to upgrade the capacity to tackle disinformation and enhance cybersecurity during elections.
- Examine the relevance of public agreements with Internet Service Providers and IT companies and international cooperation in the field of cybersecurity.
- Assess how political parties and candidates can collaborate to prevent and counter disinformation.
- Understand the importance of collaborative fact-checking, myth-busting, trust, and credibility-enhancing initiatives and the differences between a self-regulatory approach and a hybrid co-regulatory approach to online content moderation.
- Provide an overview of the phases of social media monitoring and the strategies to counter hate speech.

EMBs in different countries have responded to disinformation by collaborating with technology companies, supporting fact-checking initiatives, fostering voter education, identifying disinformation threats, and undertaking related strategic and crisis planning, coordinating with other state bodies, etc.[359]

6.1. ICT APPLICATIONS AND AI-POWERED TOOLS

There is an urgent need for EMBs to upgrade their capacity for the use of ICTs through the entire electoral cycle. The International Institute for Democracy and Electoral Assistance (IDEA),[360] has listed the ICT tools that can support the conduct of each phase:

Throughout all phases of the electoral cycle

- **Voting services web portals:** These provide the face of the EMB to the voter on the Internet, and can include voter and candidate registration services, voter records search functions, general electoral related information, instructions on how to vote, background information such as the relevant legal framework, and the official results of the election.
- **Election administration systems:** developed to support all aspects of the administration of an election, ideally such systems represent the core of an EMBs' ICT infrastructure, by building an architecture to support necessary applications and providing generic interfaces towards the other systems.[361]
- **Reporting on campaign financing:** to live up to the increasing requirements for documentation of campaign contributions,[362] reporting systems have been put in place to cope with the resulting increased workload for EMBs.

---

[359] B. Martin-Rozumiłowicz and R. Kužel, ***Social Media, Disinformation and Electoral Integrity***, IFES Working Paper. Available at: https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf. For an assessment of electoral-specific responses and related recommendations, also see https://en.unesco.org/sites/default/files/5_ecosystem_responses_aimed_at_producers_and_distributors_96_139_balancing_act_disinfo.pdf, pp.123-139.

[360] See: www.idea.int.

[361] R. Krimmer, A. Ehringfeld and M. Traxl, 2010, Evaluierungsbericht: E-Voting bei den Hochschülerinnen-und Hochschülerschaftswahlen 2009, ***Bundesministerium für Wissenschaft und Forschung***, Wien.

[362] See, among others, Organization of the American States (2012): Observing Political-Electoral Financing Systems: A Manual for OAS Electoral Observation Missions, Washington, https://www.oas.org/es/sap/deco/pubs/manuales/MOE_Manual_e.PDF; OSCE/ODIHR, 2015, Handbook for the Observation of Campaign Finance, Warsaw, https://www.osce.org/odihr/elections/135516.

- **Management of multiple voting channels:** the related mechanisms are often operated in combination with election administration systems, which help ensure the one person - one vote rule while also still enabling convenience for the voter, who can use the voting channel of their choice.
- **Inventory tracking and management:** tools for tracking the large numbers of material that EMBs need to process during an election.
- **Records management systems:** these allow for the creation, editing, publication, storage, and overall management of EMBs' records.
- **Data analysis:** election administration can use a data warehouse, which is an information system, to process, analyse and display large volumes of reporting data. Through this kind of tool, any kind of reporting in elections could be optimized and thus provide possibilities for automatic analysis.

During the Pre-electoral phase

- **Voter registration, review of electoral registers:** Supporting the voter registration process is one of the most effective ways to assist the conduct of an election by electronic means, thus increasing the level of electoral integrity.[363] This kind of support is equally sought after, as many EMBs need to build their own voter records and cannot rely on a robust population register. Ideally, such systems would not depend on Internet connectivity, as this would increase the vulnerability to attacks and potential hacks.
- **Digital ballot paper delivery:** to increase convenience, some EMBs have started to deliver ballot papers to voters, digitally via e-mail and through other remote channels.
- **Registration of election observers:** these tools enable international and domestic election observers to receive accreditation from the EMBs through an e-service.
- **Signature collection tool:** introduced by the European Union for the collection of support signatures in the context of the European Citizen Initiative, this tool could also be used for the collection of signatures in the framework of similar popular initiative instruments.
- **Party/candidate registration and ballot paper generation:** candidates and parties can also be registered digitally, which can help generate the ballot papers electronically as well as avoid data entry and typing errors, thus contributing to increasing electoral integrity.

During the Electoral Phase

- **Electronic poll books:** these can help run a smoother election by speeding up the search for voters' records, including the verification of their eligibility to vote.
- **Electronic voting machines:** these devices can support seeing-impaired voters in casting a vote independently, and also allow for the conduct of large volume elections (e.g., which need to include several levels of elections at the same time, and which involve a large number of voters). These machines are, however, heavily criticized, in particular when they do not provide any means for verification, such as the Voter-Verifiable-Paper-Audit-Trail (VVPAT).
- **Ballot paper scanners:** these tools can help count paper ballots quicker, while still retaining the paper ballot as evidence for eventual hand-recounts.
- **Ballot paper marking devices:** these electronic tools provide – together with a paper ballot scanner – the same functionality as an electronic voting machine, with the difference that the vote is stored on human-readable paper ballots. The ballot paper marking devices fill the paper ballot on behalf of the voter, upon the voter's entry of instructions (usually) on a touch-screen.
- **Internet voting systems:** For a long time, Internet voting was expected to make remote electronic voting easy, from any place in the world. However, it proved considerably challenging

[363] Z. Haque and D. Caroll, 2020, Assessing the Impact of Information and Communication Technologies on Electoral Integrity, ***Election Law Journal: Rules, Politics, and Policy***.

for most EMBs to introduce such an electronic channel, due to its inherent complexity, as it requires solving the voter identification/anonymity paradox. Estonia is currently the only country to have introduced an Internet voting channel for all elections without any restrictions for its voters.

- **Voter turnout reporting:** using this tool, every polling station can provide real-time turnout numbers throughout the day. This is relevant in relation to the interaction of EMBs with media, as voter turnout is often the only information that can be reported before the polling stations close, and is thus very interesting for media actors on election day.

In the Post-election Phase

- **Result transmission, aggregation and tabulation software:** In contrast to voter registration systems, information systems that support the transmission of results from polling stations to the regional and central levels likely have a detrimental effect on electoral integrity, largely due to the lack of time and of the possibility to switch to backup/alternate systems. Still, they continue to increase in popularity, as timely results are essential in the information society.
- **Calculation of mandates:** these useful applications help to conduct the necessary mathematical operations to allocate seats in parliament based on the election results. These systems are usually operated with no connection to the Internet, and thus, locally.
- **Systems to publish election results:** these systems work in close connection with, or are even integrated into online voter service portals, and are used to display an election's results in different levels of granularity – federal/national, regional and municipal levels, sometimes even down to polling station level, as this is increasingly requested by election observers.
- **Information for successful candidates:** integrated into the election administration system, this tool can automatically inform candidates about their success.

Moreover, EMBs can also explore the opportunities of using AI tools. These technologies have four types of capacities that can positively impact the performance and the results of governmental decision-making regarding elections:[364]

- ***Predictive analysis***, which can support policy-making by establishing linkages between different data and creating predictions.
- ***Detection***, which refers to a method that finds errors, mistakes or fraudulent behaviour, with the aim of counteracting these.
- ***Computer vision***, which refers to the possibility of analysing inputs from digital images, such as satellite images, videos or medical images.
- ***Natural language processing***, which concerns the analysis of audio and text files and can lead to automation of, for example, translation tasks.

Thus, EMBs could explore ways in which they could harness the abovementioned capacities, for instance by applying automated technologies to promote a debate that is more factually true, as shown by the examples of the Digital Democracy Room initiative in Brazil or the Avantgarde start-up in France. The automation of cognitive tasks that AI allows for can also greatly help the detection of identity fraud[365] – an area of key relevance to elections. Moreover, as election campaigning is increasingly conducted on social media platforms, EMBs could use AI-powered technology to check whether campaigning practices adhere to international standards and national legislation.

---

[364] J. Tito, 2017, ***Destination unknown: Exploring the impact of Artificial Intelligence on Government***. Available at: https://resources.centreforpublicimpact.org/production/2017/09/Destination-Unknown-AI-and-government.pdf.

[365] J. B. Bullock, 2019, Artificial intelligence, discretion, and bureaucracy, ***The American Review of Public Administration***, 49(7); T. Chen, L. Ran and X. Gao, 2019, ***AI innovation for advancing public service: The case of China's first Administrative Approval Bureau***, Paper presented at the Proceedings of the 20th Annual International Conference on Digital Government Research.

However, EMBs should also consider that applying AI technology also bears some risks:

- ***Generalization relates***, for the most part, to human error rather than to machine error. Due to their impressive performance, machine learning tools can give the impression of being flawless.[366] Nevertheless, one should always refrain from taking a machine's prediction as a certainty. Despite being very well trained and producing very accurate predictions, these tools should always remain accountable to the oversight of human beings.
- ***Underfitting*** refers to the possibility of a machine missing the underlying information that a data set contains, hence producing poor analysis and predictions. Normally, underfitting might occur as a result of applying the wrong model with respect to the issue that is being looked at.
- ***Data bias*** can create problematic results and contribute to polarization and possibly the exacerbation of pre-existing inequalities and patterns of discrimination. If biased data is used to train machine learning tools, it can produce biased outcomes.[367] Applying this to an example related to elections, if data representing certain political opinions is used, it will further reproduce the same opinions, thus negatively impacting pluralism.
- The ***black-box effect*** is related to the ability to explain and interpret, which is most pertinent to the issue of elections. It may become problematic to use AI tools given that they produce results in a so-called black box, in the sense that no individual can comprehend what was the process that led to such outcome. Some AI applications are designed in a very complete manner that does not allow external input by the human hand on the outcome of the process.

The use of the abovementioned ICT tools, as well as taking advantage of the positive potential of AI, requires EMBs to be capable of administering ICTs efficiently, effectively and, ideally, without the help of external providers. EMB staff should thus be trained to manage and deploy ICT tools and AI technology, and such capacity building is often lacking. Furthermore, there should also be a willingness, within an EMB, to reform traditional processes and implement novel technical solutions. Expectations by all electoral stakeholders increase every year but, unfortunately, the ICT administration capacity of EMBs does not, for the most part, develop at the same pace.

6.2. PUBLIC AGREEMENTS WITH INTERNET SERVICE PROVIDERS AND IT COMPANIES

Flagging, labelling, and blacklisting are all means through which content or content creators are being marked as constituting or disseminating disinformation (or as being otherwise harmful such as promoting hatred). Some online platforms allow users to flag posts as fake or false. However, to ensure transparency and communication with these companies, electoral bodies must dialogue with them and reach agreements to guarantee information accuracy during electoral processes.

BOX 22: PUBLIC AGREEMENTS BETWEEN THE MEXICAN EMB AND INTERNET INTERMEDIARIES[368]

Before the 2018 Mexican general elections, the National Electoral Institute (INE), the independent government body responsible for organizing Mexico's federal elections,

[366] J. Berryhill, K.K. Heang, R. Clogher and K. McBride, 2019, ***Hello, World: Artificial intelligence and its use in the public sector***, OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, https://doi.org/10.1787/726fd39d-en.

[367] D. Staemmler and E. Podgoršek, 2018, ***Ethics in AI: Are algorithms suppressing us?***. Available at: https://www.elsevier.com/connect/ethics-in-ai-are-algorithms-suppressing-us.

[368] Ibid.; and D. Bassante, ***Protecting the Mexican Election from Abuse***, https://about.fb.com/news/2018/06/protecting-the-mexican-election/.

signed agreements with Facebook, Google, and Twitter to identify and eliminate the spread of disinformation on the respective platforms.

With more than 3400 public positions contested, this was one of the most significant elections in the country, in which state and local elections were scheduled to take place on the same day in most Mexican states, including nine governorships.

The Memorandum of Cooperation between the INE and Facebook focused on identifying disinformation, based on Facebook and INE's shared conviction that the best way to combat it was to generate accurate, valid, and objective information. Through the MoU, Facebook committed to broadcast the presidential debates through its platform Live. It also agreed to encourage citizens to participate in the elections, including sending users reminders to vote, adding a feature with information about the elections and candidates, and informing users of where and when to vote. On its part, INE provided Facebook with real-time data on election night.

Through the public agreement between INE and Twitter, the latter committed to disseminating verified content, including via a bot that shared essential electoral information through this platform.

In turn, the agreement with Google, envisioned that the company would provide helpful information about the election through its search engine, directing voters to their polling place via Google Maps, and reminding voters about important electoral dates. Google also transmitted the presidential debates via its video-sharing platform, YouTube, and reported preliminary results on election day through INE's programme for preliminary electoral results.

BOX 23: COOPERATION AGREEMENT BETWEEN THE ORGANISATION OF AMERICAN STATES (OAS) AND FACEBOOK ON ELECTORAL INTEGRITY, HUMAN RIGHTS, AND ECONOMIC RECOVERY

The OAS and Facebook signed a cooperation agreement to work on initiatives in several areas, including electoral integrity, human rights, and economic development in the Americas in March 2021.

The agreement was signed by the Secretary-General of the OAS, Luis Almagro, and the Vice President of Global Affairs and Communications of Facebook, Nick Clegg, in an online ceremony following a working meeting between the two teams.

The agreement provides for the development and implementation of joint research projects, training programs, and dissemination of studies in areas of mutual interest. The joint initiative seeks to continue improving responses to issues such as disinformation, electoral integrity, freedom of expression or the protection of human rights defenders, in pursuit of having an increasingly plural online and offline debate.

During the meeting, the two sides discussed topics such as electoral transparency and the regional challenge to combat disinformation. Facebook's Vice President of Global Affairs and Communication explained that the company has created teams and systems to protect the integrity of elections on its platforms at key moments for democracy. Since 2017, he said, Facebook has worked on more than 200 elections around the world, many of them in Latin America.

6.3. ROLE OF POLITICAL PARTIES AND CANDIDATES TO PREVENT DISINFORMATION

Candidates and political parties have important responsibilities in connection to the increasing use of social media and AI in the context of elections. They must, among other obligations, be transparent and apply approaches that do not breach the right to privacy and data protection when reaching their voters (e.g., when using micro-targeting). Parties and candidates are also responsible for ensuring that the messaging, advertising, and content they share during the entire electoral cycle, is correct and does not itself constitute hateful content that can be criminalized depending on the legal framework.

BOX 24: ETHICAL PACT AGAINST DISINFORMATION, URUGUAY

In the run-up to the elections in 2019 in Uruguay, the six political parties represented in Parliament signed an "Ethical Pact Against Disinformation" (***Pacto ético contra la desinformación***).[369] The Pact was conceived in the framework of a campaign to combat disinformation called ***Campaña Libre de Noticias Falsas*** (that is, the "Free from False News Campaign") that was driven by the Uruguayan Press Association (***Asociación de la Prensa Uruguaya, APU***) with the support of UNESCO and UNDP, among other partners.

Under the pact, the parties committed to:

- Not generate nor promote false news or disinformation campaigns that harm political adversaries.
- Promote, among party leaders, the avoidance of actions or expressions with an aggravating tone against adversaries.
- Agree on a permanent consultation mechanism to follow up on the Pact, to quickly respond to any situation that could affect its fulfilment.

The broader "Free from False News Campaign" had three phases:

1. Signature of the Ethical Pact Against Disinformation.
2. Training of journalists and media workers on disinformation.
3. Creation of a mechanism to check false information and detect disinformation campaigns and remove them from circulation.

6.4. COLLABORATIVE FACT-CHECKINGS

As online disinformation becomes more widespread, professional journalistic standards and the values that traditional media represent – such as verification of content and publication in the public interest – are crucial.

In recent years, a myriad of collaborative fact-checking initiatives has emerged, led by media outlets and independent journalists, as well as involving other stakeholders. These efforts debunk false information, or rate content by truthfulness, among many other approaches. Some are national and others are international.

[369] See: https://www.uy.undp.org/content/uruguay/es/home/presscenter/articles/2019/04/partidos_politicos_firman_pacto_eticto_contra_desinformacion.html.

BOX 25: *VERIFICADO*, MEXICO

Verificado was launched prior to the 2018 Mexican general elections as a collaborative initiative, seeking to debunk viral and potentially harmful misinformation, fact-check politicians' claims and verify reports on the electoral process.[370] ***Verificado*** was developed by AJ+Español, the collaborative Pop-Up Newsroom and the Mexican online news site ***Animal Político***, which were joined by around 90 partners from the media, academia, and civil society sectors. The initiative was supported by the Facebook Journalism Project and Google News Initiative, Twitter, and other donors and was similar to some efforts that had been implemented in other countries, such as Electionland in the United States or CrossCheck in France.[371] The initiative was ground-breaking in terms of the number of partners involved and the support it received from Internet companies.[372]

As a part of this collaboration, Facebook provided data about the most shared stories on its accounts in Mexico to help detect possible false news. In contrast, Google provided support through its trends website to help understand what information Mexicans were looking for concerning the elections.

In addition, the materials that had been fact-checked by the whole #Verificado2018 network of partners had the «Verified» stamp when they appeared listed in Google searches, which allowed users to be certain that the concerned information was accurate. For its part, Twitter allowed the use of several tools, so that tweets with verified information and denials of false news generated by this initiative always had a preference in the platform's feed.

In the framework of the project and before the official launch of the electoral campaigns, more than 100 journalists from partner organizations came together to receive training on verification methods and define metrics, roles, functions, and workflows collaboratively. During the four months prior to and during the elections, 400 articles, more than 100 visuals, and educational material on verification were produced and disseminated via social media and ***Verificado's*** website. The team also fact-checked in real-time the three presidential debates held during the electoral process. Facebook Live videos were streamed during election day and those preceding it, raising awareness about the initiative, and broadcasting directly from the National Electoral Institute.

The initiative invited the public to make requests for the verification of claims, images, or videos, by using a specific hashtag (#Quieroqueverifiquen, which means "I want you to verify") on Facebook and Twitter. There was also a WhatsApp feature available, through which users could choose to receive compiled debunks.

---

[370] See: https://verificado.mx/metodologia/.

[371] See: https://wan-ifra.org/2019/11/verificado-2018-fighting-misinformation-collaboratively/.

[372] K. Bontcheva and J. Posetti (eds.), 2020, ***Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression. Broadband Commission research report in 'Freedom of Expression and Addressing Disinformation on the Internet***'. Available at: https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf.

BOX 26: *CHEQUEADO*, ARGENTINA

During the 2019 elections in Argentina, more than 100 media actors and technology companies collaborated in the ***Reverso***[373] initiative, which was supported and coordinated by fact-checking organization ***Chequeado***, AFP Factual, First Draft and Pop-Up Newsroom.[374]

All partners simultaneously published the debunks produced by ***Reverso***, to maximize their visibility. During the 6 months that the electoral campaign spanned over, ***Reverso*** developed 180 articles and 30 videos. Content was monitored on Facebook, Instagram, Twitter, WhatsApp, YouTube, the crowdsourcing platform *Chequeo Colectivo* (which is managed by ***Chequeado***), and other platforms. Reportedly fake audio files featuring candidates and shared via private messaging apps were also verified,[375] in cooperation with BlackBox.[376] The ***Reverso*** team also notified state authorities about existing information gaps, inviting them to publish official communications as needed.

An important lesson learned from the initiative is that while fact-checking efforts of this kind do not necessarily change people's minds about causes or issues they are profoundly convinced about, they do have the potential to impact individuals' behaviour by deterring them from resharing information that has been demonstrated to be false (fearing how they could be perceived if they did), thus contributing to diminishing the spread of disinformation.[377]

BOX 27: EXAMPLES OF TRUST AND CREDIBILITY ENHANCING INITIATIVES

In recent years a variety of efforts has emerged focused on building trust in news media, as well as on rating/labelling them according to their credibility. Some examples:

The **Ethical Journalism Network** is a worldwide coalition that brings together over 70 groups of journalists, editors, media owners and media support initiatives, aiming to promote accountable journalism, through capacity-building, research and advice to policy-makers and human rights advocacy groups.[378]

The **Journalism Trust Initiative** is a collaborative standard-setting process that seeks to support journalism and counter disinformation. It does so by facilitating the development of trust and transparency standards to be adhered upon and implemented by media. It was created by Reporters without Borders (RSF), Agence France Presse and the European Broadcasting Union in April 2018.[379]

---

[373] See: https://reversoar.com/.

[374] See: K. Bontcheva and J. Posetti (eds.), 2020, https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf.

[375] Ibid.

[376] See: https://blackvox.com.ar.

[377] According to research focused on the impact of **Chequeado** and **Reverso**, which was referred to by these initiatives' Director, Laura Zommer, during a session organized by UNESCO and UNDP at the 2020 Internet Governance Forum. Available at: https://en.unesco.org/news/confronting-disinformation-electoral-process-call-coordinated-action.

[378] See: https://ethicaljournalismnetwork.org/.

[379] See: www.jti-rsf.org.

Newsguard reviews and rates online news sites in terms of their trustworthiness, through a web browser extension that lets customers know if a news website that they are navigating is reliable. This US-based commercial initiative employs experienced journalists as analysts, who undertake the review and rating process according to criteria such as whether a website regularly publishes content that is false, collects information and reports on it fairly and accurately, is transparent about conflicts of interest, discloses its ownership and financing sources, clearly labels ads, provides the names of content creators, corrects or clarifies reporting errors publicly, avoids deceptive headlines, etc. Newsguard supports advertisers by warning them to avoid the publication of their ads on websites that could negatively impact their brands.

Credder is a community-driven initiative which permits journalists and the public to review and rate the credibility of articles, their authors and the sources they cite, generating statistics based on this input. The assessments provide guidance to the consumers of information, orienting them toward content of a stronger quality and strengthening their discernment as it concerns lower quality information.[380]

6.5. USING AI TO COUNTER DISINFORMATION

The implementation of innovative research that applies AI and other technological developments also has a significant role to play in responding to the identified emerging challenges all along the electoral cycle. The first case study below presents an example of an impactful applied social research tool to monitor and reinforce public debate, as well as to curb disinformation in an electoral context. In turn, the second case study featured in this section focuses on a start-up that aims to supports democracy, promote voter engagement and counter disinformation through the ethical use of algorithms.

BOX 28: THE DIGITAL DEMOCRACY ROOM, 2018 PRESIDENTIAL ELECTION IN BRAZIL

An often-cited example of concerns related to the bot-powered spread of disinformation via messenger services is the 2018 Presidential election in Brazil.[381] WhatsApp served as a platform to micro-target the electorate in groups, focusing on certain political issues that they might be interested in,[382] through "bulk messages" sent automatically and at a large scale.[383]

In the year of the elections, the Brazilian think-tank and higher education institution Fundação Getulio Vargas (FGV) launched, through its Department of Public Policy, a new initiative called Digital Democracy Room (DDR). The DDR monitored public debates in the digital environment during elections and fights the spread of online disinformation and misinformation that could undermine the integrity of the political and electoral

[380] K. Bontcheva and J. Posetti (eds.), 2020; https://credder.com/.

[381] E. Bracho-Polanco, 2019, How Jair Bolsonaro Used 'Fake News' to Win Power, ***The Conversation***, Jan, 8.

[382] M. Magenta, J. Gragnani and F. Souza, 2018, How WhatsApp is being abused in Brazil's elections, ***BBC News Brasil***.

[383] Ibid.; M. A. Ruediger, A. S. Grassi and L. C. Guimarães, 2020, ***Digital Democracy Room Brazil/Interviewer: R. Krimmer & N. Licht***.

process.[384] DDR followed an approach to studying democracy and the electoral context, observing the real-time political debate that was taking place on social media platforms. The approach combined a set of methodologies, drawing from the fields of linguistics, sociology, communication, statistics and information technology. It integrated these domains in order to be able to extract a high volume of data from the Internet and organize it, as well as to undertake a scientifically rigorous social and political analysis.[385]

The DDR published reports analysing trends, with the aim of strengthening democratic institutions.[386]

BOX 29: *AVANTGARDE*, FRANCE[387]

Avantgarde is a French start-up that uses AI to fight disinformation and propaganda, including those that are fuelled by computational amplification. It uses machine learning technology to flag bots that spread false information, notifying users about content that features it. It aims to support democracy and civic engagement by also educating voters and helping them access a diversity of political content, breaking echo chambers and allowing them to make up their own minds on a variety of issues.

Further, **Avantgarde** has cross-disciplinary expert teams that work with candidates and social movements to help them deliver "better AI-powered campaigns in an ethical and legitimate way".[388] It does so by combining data on attitudes, behaviour, personality and networks, harnessing algorithms that promote a meaningful and individualized relationship with citizens, so "that personalised political ads always serve the voters and help them be more informed, rather than undermining their interests".

Avantgarde is based on the premise of responsible use of machine learning, deploying bots that "disclose themselves as bots and serve the public good".[389]

6.6. SOCIAL MEDIA MONITORING

Social media monitoring is an essential tool within EMBs' strategic and crisis planning. It allows for the anticipation of disinformation threats and for preparing appropriate responses. It can also offer a solid basis for advocacy efforts, both in the short-term, medium-term and long-term, by promoting improved regulation.

6.6.1. THE PHASES OF SOCIAL MEDIA MONITORING

Before monitoring, it is necessary to undertake a capacity assessment and consider which social

[384] M. A. Ruediger, A. S. Grassi and L. C. Guimarães, 2020.

[385] See: https://observademocraciadigital.org/en/metodology/.

[386] M. A. Ruediger, A. S. Grassi and L. C. Guimarães, 2020.

[387] See: https://twitter.com/avantanalytics; https://www.weforum.org/agenda/2017/05/macronleaks-have-changed-political-campaigning-why-macron-succeeded-and-clinton-failed ; https://techsgood.org/vyacheslav-polonski-solve-our-technology-problem-with-humanity-aebf7ae5452.

[388] See: https://techsgood.org/vyacheslav-polonski-solve-our-technology-problem-with-humanity-aebf7ae5452.

[389] See: https://twitter.com/avantanalytics.

media platforms are relevant for the exercise and if there are any rules and regulations that should be taken into consideration, including the legal obligations for the platforms.

The monitoring depends on the objectives pursued, and determining its scope also entails deciding which platforms to monitor. Suppose the main aim is to know the impact of social media on the integrity of elections and what kind of information is available to voters. In that case, an essential step will thus be to identify the social media platforms that are most popular as sources of political information. The availability of human and financial resources is another factor that affects decisions on the monitoring's scope.

It is also important to make a baseline assessment of the country's social media situation and explore whether other social media monitoring efforts are being planned or implemented by CSOs, media regulatory entities, EMBs, media outlets, academics, or international observers. While it is positive to avoid duplication, the existence of social media monitoring exercises occurring in parallel is not problematic, as long as the methodology and goals of each are transparent. It is, in any case, relevant to facilitate synergy and collaboration between approaches to maximize their impact. There are distinctions between social media when content is used in a public space, which affords a degree of monitoring when content (including advertising) is more micro-tailored and distributed, and when content is in private groups (including in social messaging).

BOX 30: STRATEGIES TO COUNTER HATE SPEECH

There are a number of strategies that EMBs and electoral stakeholders can employ to counter hate all along the electoral cycle, including:

Engaging with all key actors

Displaying model behaviour, both by not engaging in hate speech or discrimination as an institution, and by not allowing any of its members or personnel to do so.

Speaking out against discrimination and hatred, to raise public awareness of hate speech and its impact.

Creating or expanding spaces for pluralistic public dialogue.

Investing in technology and specially trained human resources.

Putting in place processes and mechanisms for monitoring, data gathering and reporting.

Undertaking security planning in order to mitigate the risk of electoral violence or incitement to it.

Building capacities among electoral stakeholders on issues pertaining to human rights, voting rights, non-discrimination, gender equality, protected and prohibited forms of expression, hate speech and incitement of hatred, and relevant national legislation and international standards.

Implementing awareness-raising and voter education efforts – advocating for improvements in the national legal framework if it is inadequate for tackling hate speech during electoral periods, too challenging to implement, or not aligned with international standards and best practices.[390]

[390] V. Mohan and C. Barnes, 2018. Available at: https://www.ifes.org/sites/default/files/2017_ifes_countering_hate_speech_white_paper_final.pdf.

BOX 31: COUNTERING HATE SPEECH IN ELECTIONS IN INDONESIA

In recent years, elections in Indonesia have seen intercommunal and sectarian incitement violence flare-up, including via fake stories based on actual events that were utilized to fuel tensions for partisan purposes. The fact that many Indonesians are avid and active uses of social networks[391] contributed to the spread of hate speech and the eruption of violence, in a country with multiple religious, racial and ethnic groups.[392] This called for different actors to coordinate efforts to strategically address hate speech.

Ahead of the regional elections held in the country in 2015, for example, the Elections Supervisory Agency and the General Elections Commission collaborated with the Ministry of Communications and the National Police to strengthen monitoring and countering of hate speech. While also referring to the government's regulatory, legal and technology-based approaches, the country's Minister for Communications and Information publicly supported awareness-raising campaigns in particular, as an "even more effective way to stop the spread of discriminatory and hateful speech". He urged prominent figures and civil society to become involved in sensitization actions.[393]

In the lead-up to the 2018 and 2019 elections in the country, Bawaslu (an official entity established to monitor elections), the General Commission of Indonesia and the Ministry of Communications and Information Technology signed a Memorandum of Understanding to work together against disinformation and hate speech. This included joint actions to monitor false news and incitement to intercommunal violence in the social media accounts of political parties and candidates.[394]

BOX 32: THE DIGITAL DISINFORMATION COMPLAINTS COMMITTEE FOR SOUTH AFRICAN ELECTORAL COMMISSION [395]

Ahead of the 2019 elections in South Africa, the CSO Media Monitoring Africa[396] piloted an innovative multi-stakeholder approach to countering hate speech and disinformation in electoral contexts: a complaints committee that functioned based on the crowdsourced gathering of complaints about online content. The Digital Disinformation Complaints Committee was set up with the aim of advising the Independent Electoral Commission (IEC), given public concerns about the impact of hate speech and disinformation in the run-up to the elections. The IEC, the South African National Editors Forum and the Press Council of South Africa endorsed this initiative.

---

[391] With a penetration rate of over 88 percent, YouTube was the most used social network in Indonesia according to data for the 3rd quarter of 2019. However, all the other widely known social media platforms such as WhatsApp and Facebook also enjoy a high penetration rate there, making the country one of the largest social media markets in the world. Source: https://www.statista.com/statistics/284437/indonesia-social-network-penetration/.

[392] See: https://www.ifes.org/news/countering-communal-incitement-and-hate-speech-indonesia.

[393] V. Muntarbhorn, ***Study on the Prohibition of Incitement to National, Racial or Religious Hatred: Lessons from the Asia Pacific Region***, undated. As cited in IFES White Paper.

[394] See: https://www.ifes.org/news/countering-communal-incitement-and-hate-speech-indonesia.

[395] T. McGonagle, M. Bednarski, M. Francese Coutinho and A. Zimin, 2019, ***Elections and media in digital times***, In Focus edition of the World Trends in Freedom of Expression and Media Development, UNESCO, Paris.

[396] See: https://www.mediamonitoringafrica.org.

As part of this effort, a REAL411 platform[397] was created. This platform allowed individuals to report cases of disinformation and hate speech, as well as follow the status of these complaints, including the actions taken in response. The platform was connected to the Directorate of Electoral Offences. The Digital Disinformation Complaints Committee – including specialists in media law and representatives of social media platforms and online media outlets – reviewed the complaints received and issued recommendations for possible action on the part of the Electoral Commission (e.g., referring the case for criminal or civil legal action, requesting the content's removal by social media platforms, producing press releases to warn the public and correct false information).

The initiative received support from platforms like Facebook, Google, and Twitter, which also undertook content moderation on instances of disinformation. Related voter education campaigns contributed to sensitizing the electorate on hate speech and disinformation.

In turn, the decisions on disinformation campaigns were guided by a Draft Code on Disinformation during Elections that was developed in addition to the already existing South African Electoral Code of Conduct. Under this draft code on disinformation, registered political parties were requested to upload their official online advertisements on an online repository that was piloted during the 2019 electoral process, which the public could use to compare legitimate political advertising and instances of disinformation.[398] More recently, the initiative has been adapted to tackle disinformation related to the COVID-19 pandemic.[399]

6.6.2. DATA SCRAPING AND CODING

One of the most challenging aspects of social media monitoring is accessing data, as policies in this area are continuously changing. Social media companies are very cautious when it comes to providing access to data, and their obligations vary from country to country and region to region. It is possible to request full access to a platforms' API (Application Programming Interface Access), yet obtaining it varies across platforms.[400] Twitter is the most open platform in this regard, as it provides more direct access to researchers via its API (or through Gnip, an API aggregator company owned by Twitter).

Being more widely used, Facebook is oftentimes more relevant for monitoring purposes, yet the Meta shares data only with specific partners through Crowdtangle. This tool gives access to Facebook and Instagram data, including the number of interactions related to a post, link, video, and were the most shared. It facilitates the identification of viral content and makes it possible to check whether it features false information. Nevertheless, researchers do not enjoy widespread access to this tool for advanced media monitoring yet. Facebook has also gone as far as taking legal action against a university for recruiting volunteers to share data about the advertising feeds to which they are exposed.[401]

---

[397] See: https://www.real411.org.za/. The inclusion of "411" in the platform's name alludes to this number standing for "information" in internet slang; K. Bontcheva and J. Posetti (eds.), 2020.

[398] See: http://www.padre.org.za.

[399] K. Bontcheva and J. Posetti (eds.), 2020.

[400] Even if API access is available, obtaining it can take a long time, and incur large administrative and bureaucratic hassles, constantly changing rules of access and data handling requires an intense and constant maintenance effort. For more information, see also: https://www.wahlbeobachtung.org/en/social-media-monitoring-results-about-2019-austrian-snap-elections-published/.

[401] Facebook Seeks Shutdown of NYU Research Project Into Political Ad Targeting. ***Wall Street Journal***, 23 October 2020, https://www.wsj.com/articles/facebook-seeks-shutdown-of-nyu-research-project-into-political-ad-

Other commercial tools available for data scraping and analysis are Newswhip, Buzzsumo, Vlsibrain, Sysmos, Talkwalker, and BrandWatch. Which tool to select will depend on the goals of the monitoring to ensure that it allows access to the needed data. Newswhip, for instance, permits an analysis of the content on a specific topic that goes viral on Facebook, Instagram, YouTube, Twitter, and other websites. It can predict a link's future popularity and, thus, its impact based on past interactions. Other important elements to decide on are the unit of analysis (e.g., a single post) and the system of coding. Enlisting users to assist in monitoring, especially on closed channels such as private Facebook groups and social messaging, can be a way to extrapolate insights even though the resulting data may not be generalizable.

6.6.3. DATA ANALYSIS

Analysis of the data is both quantitative and qualitative based on previously identified indicators and parameters. Depending on the chosen criteria, it can focus on the number of posts and interactions like the total number of reactions, comments, and shares. When evaluating the activity of political actors on Facebook, it should be noted that some parties and candidates promote their posts using paid advertising ('boosting'), which influences a post's reach and may therefore amplify the impression of its "success".

There are several aspects that could be monitored in relation to social media's role during an election, including the collection and analysis of data regarding online campaigning of parties and candidates; election advertising; divisive narratives and disinformation campaigns; and dangerous or hateful speech.

When it comes to the length of the monitoring, this very much depends on the available human and financial resources. It is a good practice to cover a minimum period of at least two weeks but the longer the monitoring period, the better validity of the data.[402]

Another aspect that social media monitoring could look at is the method/tactics for dissemination, such as information operations, targeted attacks against vulnerable groups and voter suppression.

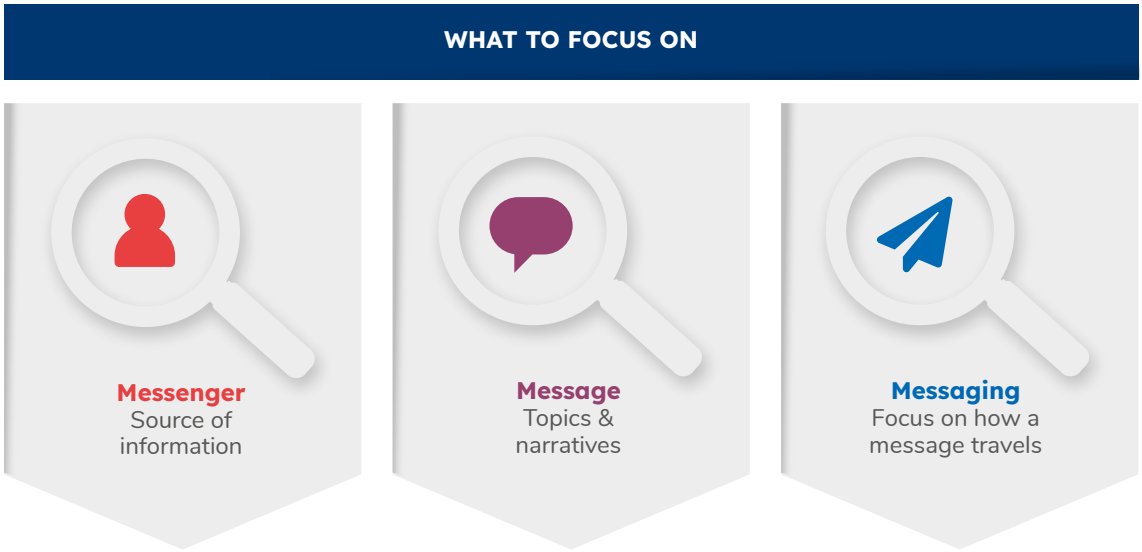


FIGURE 10: THE FOCUS OF SOCIAL MEDIA MONITORING

targeting-11603488533.

[402] See: https://rm.coe.int/monitoring-of-media-coverage-of-elections-toolkit-for-civil-society-or/1680a06bc6.

Monitoring should look at:

- **Messenger:** The source of information, that is, the actor who disseminates the message. It could be a candidate, a political party, a media outlet that promotes a story through social media, an "influencer" (someone with a significant number of followers and supporters on social media), etc.
- **Message:** The topics and narratives that are spread by the messengers. Content analysis can examine if these are used by specific actors to disseminate disinformation, confusion, etc. Social media monitoring selects certain deliberately expressed perspectives on a particular issue and looks at the frequency in which they appear on social media posts. The choice of which narratives to focus on depends on the political discourse in each country. Monitors should be ready to add new narratives to take note of, if they start to be significantly debated as the electoral process unfolds.
- **Messaging:** The manner in which the message is disseminated on social media platforms. For example, whether it is amplified by bots or trolls, or if it is shared as sponsored content, and thus has its visibility boosted.
- **Amplification:** an important area to examine is prioritisation of content received by voters, both in their "news" feeds and in the advertising presented to them. There will normally be variation due to personalisation algorithms, and yet at the same time, patterns may emerge as to curation engineering that privileges some kinds of content on the "enragement = engagement" rationale of keeping users glued to a given platform and yielding ever-increasing amounts of data as they are drawn down a "rabbit-hole" of increasing intensity.[403] Methodologically, there are challenges for suitable sampling to monitor this kind of content play, but tracking along the lines of Facebook's look-alike audience categorisations can be explored.[404]

6.6.4. REPORTING

The number of interim reports to be produced during the monitoring effort will depend on its length. In turn, the final report should reflect a comprehensive analysis that includes thorough information about the trends identified, as well as the presentation of monitoring results (in the form of charts and tables), integrating both a quantitative and a qualitative approach. The final report should also outline recommendations, including improvements in different areas, for example, in relation to the legal framework and other relevant dimensions.

BOX 33: SOCIAL MEDIA MONITORING OF THE 2019 NIGERIAN GENERAL ELECTION

The European Union deployed for the first time in the framework of an Election Observation Mission (EOM), a social media analyst to track the influence of digital forms of communication on the 2019 Elections in Nigeria. According to the 2019 DataReportal Digital report resulting from this initiative, Facebook and WhatsApp were the two most popular platforms in the country, followed by Instagram and YouTube.[405]

The EU EOM's social media monitoring identified several tactics used to discredit opponents, mislead voters, and cast doubt on the electoral process. While by the time of the elections Facebook had launched its Ad Library, the latter was not yet available for

[403] K. Roose, https://www.nytimes.com/column/rabbit-hole.

[404] See: https://www.facebook.com/business/help/164749007013531.

[405] See: https://www.slideshare.net/DataReportal/digital-2019-nigeria-january-2019-v01.

Nigeria. In its final report, the EU EOM concluded that, while offering new opportunities for the public to scrutinize the electoral process and for campaigning (notably for opposition candidates and parties, in view of potential state media bias), social networks also facilitated the lack of transparency in political advertising, and the creation of a 'fake' appearance of widespread support for certain candidates.

The monitoring served to identify several negative trends, such as manipulation through astroturfing (to simulate mass support), the use of bots on Twitter to amplify partisan messages, enlistment of social media influencers, deployment of disinformation against political opponents, as well as to distort the public's perception of the electoral process. It also detected the utilization of unrelated or fake audio or footage, uncontextualized information, fake opinion polls, and the dissemination of false results once the election was over.[406]

BOX 34: SOCIAL MEDIA MONITORING, 2019 ELECTIONS TO THE EUROPEAN PARLIAMENT [407]

From 1 April to 15 May 2019, ahead of the European Parliament elections, MEMO 98, a Slovakian non-profit specialist media monitoring organization, carried out the monitoring of the public Facebook accounts of parties in the Czechia, Hungary, Poland, and Slovakia that were running in the European elections.

The key objectives were:

- To assess social media's role during elections.
- To measure the potential impact on electoral integrity of the messages
- To evaluate public trust and confidence in the process.
- To identify the trend of topics addressed by political parties on their accounts in the period leading up to the elections.

The MEMO 98 monitoring followed 48 political parties in the four countries based on the parties' popularity ratings and status (i.e., parliamentary, and non-parliamentary). It monitored Facebook because it was among the most widely used social media platforms in all four countries, as per the Digital 2018 Global Overview Reports reports published by Datareportal.[408]

During the first phase of the monitoring, MEMO98 used Netvizz, a digital tool that exctracted data from different sections of the Facebook platform (groups, pages, search) for research purposes. File outputs can be easily analysed in standard software.[409]

The monitoring experts studied the posts' content, coding it according to a list of narratives, ranging from more global and cross-regional themes to more local issues specific to each country.

---

[406] EU Election Observation Report, Nigeria 2019. Available at: http://www.eods.eu/library/nigeria_2019_eu_eom_final_report-web.pdf.

[407] See: http://memo98.sk/uploads/content_galleries/source/memo/ep-elections-2019/fb-monitoring-ep-elections_shorter-version_fin.pdf.

[408] For more information and statistics of social media use in all four countries, see also the Global Digital report 2018. Available at: Digital in 2018: World's internet users pass the 4 billion mark - We Are Social UK.

[409] Netvizz was an open-source tool written and maintained by Bernhard Rieder, Associate Professor in Media Studies at the University of Amsterdam, The Netherlands, and researcher with the Digital Methods Initiative. The tool is no longer available given Facebook's decision.

The monitoring also took into consideration narratives that had been used in instances of foreign interference in previous elections in other countries to diminish citizens' trust in their democratic institutions. It analysed if such narratives were used by any political parties running in the elections to spread disinformation and confusion, and if they had a disruptive impact on electoral integrity.

From a more general perspective, the monitoring experts examined to what extent parties and candidates used Facebook for campaigning, mobilizing voters, and voter education purposes. It focused on possible signs of dangerous speech and inflammatory language in the posts observed, as well as efforts to discredit political opponents. Finally, it also tried to assess whether the posts focused on important public policy matters or mainly aimed to get attention by referring to scandals, conspiracies, and myths.

MEMO 98 monitoring determined the focus issues for each monitored party and which topics/narratives generated the highest engagement (comments, shares, and reactions). It revealed that political parties focused more on domestic political scenes than EU-related topics and privileged some divisive and attention-grabbing issues, such as migration. Nonetheless, many other parties ran campaigns focused on positive aspects and highlighted the benefits of European integration. Social media monitoring did not show a disinformation campaign of the magnitude of others seen during recent elections worldwide.

6.6.5. MONITORING CHALLENGES

The exercise of social media monitoring during the electoral process is an enormously helpful tool, but it also presents challenges. To start with, although it is possible to monitor certain elements related to electoral campaigns (e.g., hate speech content), exhaustive monitoring of social media networks is impossible due to the speed, reach, and volume of the content produced, along with the unlimited number of existing pages, accounts, and websites, as well as the opacity of the social media companies.

The lack of access to the data of crucial social platforms restricts the extent of the analysis too. As many dimensions of electoral campaigning through social media remain unregulated, the information retrieved by focusing on legal compliance may be limited. The multi-media nature of the content disseminated (text, video, audios etc.) also makes social media monitoring particularly complex. Additionally, the closed nature of conversations taking place in private groups limits the scope of monitoring that it is possible to undertake. At the same time, the monitoring of closed groups presents serious concerns regarding privacy or the anonymity of users' accounts.

BOX 35: MONITORING ELECTIONS ON SOCIAL MEDIA IN TUNISIA

Since the overturn of Tunisia's former president Ben Ali, during the Arab Spring in 2011, in which social media platforms and particularly Facebook played a vital role, social media platforms have become a vital tool to host public and semi-public political discourse in the country.[410] In 2019, with a total of 7.4 million users, Facebook, was the most used social network by Tunisians to gather electoral information.[411]

[410] A. Mhenni and H. d. Baillenx, 2019, ***Monitoring of electoral campaigning on social media – Tunisia***, ATIDE & DRI, Tunisia.

[411] Ibid.

Nevertheless, in recent elections social media platforms, generally, have proved to also bear risks to electoral integrity.[412] The fast spread of disinformation and intimidation tactics toward electoral stakeholders has negatively affected Tunisia's electoral integrity.

In 2019 election, the Association Tunisienne pour l'Intégrité et la Démocratie des Elections (ATIDE) partnered up with Democracy Reporting International (DRI). ATIDE is a Tunisian entity that monitors electoral activity and with DRI's support extended its observations to social media platforms.

The cooperation of ATIDE and DRI implemented certain analysis strategies applying AI-powered systems and data analytics.[413] For example, they mapped the political content creators, analysed which platforms would be needed to be monitored, what information should be collected, and what means were available to collect the necessary data.

Via semantic detection techniques, much of the floating political content was able to be identified.

In their analysis, ATIDE and DRI identified many pages on Facebook with high political engagement of which more than 40% were not transparent as to their political affiliation, ownership, or target. Most of these pages described themselves as entertainment or satirical sites without any political affiliation and yet engaged with high frequency in the content distribution of sponsored messages (over two thirds of all political messages from non-media-related pages were produced by these pages). Furthermore, ATIDE and DRI's analysis detected that often unofficial political pages did not adhere to standards set out for election campaigns. In addition, the analysis detected that identical political content was systematically posted from different sources which created the assumption of a connected network or a centrally administrative entity (some of which changed their political affiliation in the course of campaigning).

Another insight was that Tunisia's already existing monitoring entities, the Instance Supérieure Indépendante pour les Élections (ISIE) and the High Authority for Audiovisual Communication (HAICA) could improve their capacity in detecting content that bypassed their monitoring systems.

Regarding investigating ads and placement containing disinformation, many were erased after the election and Facebook would have needed to provide ads libraries to investigate these ads further.

6.7. SELF-REGULATORY APPROACH TO ONLINE CONTENT MODERATION

Given the growing consensus on the need for increased transparency in social media platforms' content moderation,[414] the international CSO ARTICLE 19 has proposed a model focused on creating Social Media Councils (SMCs). As mentioned previously in this Guide, this approach was discussed with social media companies and received the endorsement of David Kaye during his mandate as UN Special Rapporteur on Freedom of Opinion and Expression.[415]

---

[412] Ibid.

[413] A. Mhenni and H. d. Baillenx, 2019.

[414] See: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/gdpiart_19_smc_conference_report_wip_2019-05-12_final_1.pdf.

[415] See: https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

The proposed mechanism for self-regulation has a multi-stakeholder approach that aims "to provide an open, transparent, accountable, and participatory forum to address content moderation issues on social media platforms based on international standards on human rights".[416]

While the decisions and recommendations of these councils would not be legally binding, companies and digital platforms would commit to abide by them in good faith and of their own accord. The SMCs would review the platforms' content moderation decisions according to human rights-based principles. They would propose non-pecuniary remedies such as the right of reply, the issuing of an apology in cases where the content was removed by error, the publication of a decision, the re-uploading of the removed content, etc.

The SMCs approach is only for online content moderation, not for personal data collection, online advertising, or taxation issues, which its proponents suggest should be tackled through alternative mechanisms. At the same time, the SMCs would not review governments' requests for the removal of content on the platforms, as this kind of review should be carried out by an independent judicial body, in line with the national and international legal frameworks.

The main characteristics of the SMCs are:

- **Independence.** They are independent of the State, businesses, and other special interests.
- **Inclusiveness.** They must rely on inclusive, broad, and transparent public consultation process.
- **Democratic.** Membership selection and decision-making must be transparent and democratic.
- **Diverse.** Broad representation, reflecting the diversity of society in its composition.
- **Follow clear rules.** They should include a strong complaints mechanism, follows clear rules to determine whether there was a breach of standards in the reviewed cases, and have the power to impose moral sanctions only.
- **Transparent.** They must act in the service of public interest, in a transparent and accountable way. UNESCO has explored with Article 19 building coalitions to monitor social media disinformation in three pilot countries, as part of its Social Media for Peace project.

6.8. THE HYBRID CO-REGULATORY APPROACH

In the UNESCO-commissioned paper "Social media and elections", Andrew Puddephatt recommends adopting a hybrid system of co-regulation concerning social media during electoral processes. In its framework, the EMBs would not impose detailed prescriptive requirements on social media platforms, but rather set the expected outcomes, develop a code of practice for social media companies (in consultation with these companies, as well as with political parties, and the wider public), review, and make public "the way industry has, or has not, met the standard".[417]

This model understands that the role of an EMB is assuring the outcome of free and fair elections and facilitating the electoral process. Nevertheless, EMB should not decide what parties or individuals can and cannot express during the electoral cycle, but only set the parameters that apply to all participants, including the government.

EMBs should:

- Understand the domestic context, including the laws and regulations that apply, their impact on the electoral process and which other entities are concerned. The EMB should liaise with these bodies, such as the data protection office or ombudsperson, the security service, the police, etc.
- Consider international human rights standards and their application to social media platforms,

[416] Article 19, June 2019, pp. 7-8.

[417] See: A. Puddephatt, 2019, ***Social media and elections***, UNESCO, Paris, p. 25. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000370634.

including the related complexities and risks.

- Define the scope of the online platforms whose practices can come under its jurisdiction during an electoral process (in this regard, a focus on the largest and most popular ones is suggested).
- Avoid unintended harms in terms of unduly restricting political campaigning.
- Focus on achieving outcomes by setting goals for Internet intermediaries, in relation to what they are expected to do and how to ensure a free and fair election. In this respect, the EMB should open a dialogue with Internet companies and establish a partnership with them, based on a discussion on the possible concerns that may emerge and on how the companies will respond while also safeguarding citizens' rights. The partnership should establish accountability mechanisms through which illegal conduct on the platforms can be addressed by the EMB (e.g. making it possible for it to impose fines on political parties that do not comply with electoral funding legislation), as well as transparency requirements on companies (the paper recommends that, if social media platforms are aware that they are permitting disinformation to be presented as news during the electoral period, they should face proportional financial penalties or prosecution after the election).
- Develop a Code of Practice setting the standards for a free and fair election, in the run-up to it and in consultation with political parties, private sector actors, civil society and the public. The Code should cover areas such as the companies' commitments and procedural safeguards in relation to content removal, management of political advertising, their establishment of an accelerated complaints procedure, their regular reporting on measures against misinformation, disinformation and the misrepresentation of identities, as well as transparency regarding the use of their platforms.
- Consider creating a Rapid Alert system to issue warnings regarding disinformation in real-time.
- Encourage auditing by academia, journalists and CSOs on the use of social media algorithms to give visibility to certain stories during the electoral process.
- Implement a public information campaign creating awareness during elections and about the risks of relying solely on social media.
- Coordinate with the Data Protection Authority to ensure the effective protection of personal data during the electoral campaign.
- Invite social media platforms to explain how they plan to achieve the set outcomes and address the key concerns regarding their role in elections and require them to report – after the election – on how they have fulfilled their commitments in connection with the Code of Practice. The EMB should publish a report based on this information and other relevant aspects of the experience.

6.9. GUIDANCE CONCERNING THE REGULATION OF ONLINE CONTENT MODERATION

The Global Network, an Internet-industry association, following a human-rights based analysis of more than 20 governmental regulations, identified several recommendations for EMBs to tackle diverse forms of digital harm:[418]

In relation to the principle of Legality:

- The process for developing laws and regulations should follow an open, participatory approach.
- Independent bodies that are given power and discretion in terms of rulemaking should be subject to strong oversight and accountability mechanisms.

[418] For a full and detailed list of GNI's recommendations, see: https://globalnetworkinitiative.org/wp-content/uploads/2020/10/CRPB-Appendix-A-Recs.pdf.

- Legislation should define clearly and narrowly what is prohibited and who can be held liable for not enforcing such prohibition. It should also set clear expectations regarding companies' duties concerning reports of illegal content.
- Laws should require transparency, oversight, and adequate remedy, so that those implementing them does not have undeterred discretion for restricting freedom of expression.

Regarding the principle of Legitimacy:

- Laws should be formulated with precision and establish clear criteria for permissible limitations which are those only permitted exhaustively by international law. The decisions on whether those criteria are met or not should be taken by a judge.
- Any prohibition of content should be aligned with the legitimate purposes outlined in Article 19 (3) of the ICCPR.
- Controversial or offensive content cannot be prohibited only on account of being uncomfortable for certain people.

In relation to the standard of Necessity:

- Restrictions should be based on the establishment of a clear link between a particular expression and an alleged threat.
- Laws should be appropriate to achieve their protective goals, yet in the least intrusive manner. Inclusive public consultations are recommended to achieve this aim.
- Consideration should be given to how the requisites set in law may impact diverse types of firms (e.g., how they may affect start-ups and small companies), accommodating a range of different business models and capacities, and considering possible effects on competition policy.
- Laws should provide clear guidance in terms of the nature of the content and the circumstances that call for fast or significant action.
- The principles of transparency, due process, remedy, and other traditional rule of law concepts should guide the development of standards for online content moderation.
- Legislation should give room for experimentation and flexibility in approaches, while also including safeguards against intentional misuse and unwanted consequences of content removal actions.
- There should be remedial mechanisms to which users can resort when their content is restricted, to avoid encouraging self-censorship and over-removal.
- Laws should allow for periodic reviews, so that they remain up to date in connection to the fast evolution of norms and technologies.

To protect Privacy:

- While enhancing accountability on the part of those who violate the law, those behind the development of regulatory efforts should also find ways to ensure that privacy protection is strengthened for all.
- Regulation efforts should acknowledge that anonymity and pseudo-anonymity can help shield vulnerable users from harassment, as well as the value of encryption to protect users, ICT services and the ICT ecosystem.
- Authorities should be mandated to meet due process obligations, as well as to provide required evidence before requesting access to sensitive user data.

BOX 36: RECOMMENDATIONS OF THE UK ELECTORAL COMMISSION ON DIGITAL CAMPAIGNING

Based on research focused on financial regulation and financial campaigning, in 2018 the British Electoral Commission published a report outlining a series of recommendations for the country's governments, legislatures, social media companies and campaigners.[419]

Among others, the report suggests that each of the UK's governments and legislatures should change legislation so that digital material used for electoral campaigns is required to include an imprint making clear "who is behind the campaign and who created it", as well as to change reporting rules so that campaigners distinguish diverse types of spending and provide more information about resources spent on digital campaigns. In addition, they should clarify that spending by foreign organizations or bodies on election and referendum campaigns should not be allowed. The Electoral Commission also requested governments and legislatures to increase the fines that it can impose on campaigners who do not abide by the rules and enhance its powers to access information outside of an investigation.

Likewise, the report recommends requiring campaigners to provide more detailed and meaningful invoices from their suppliers.

In turn, it also calls for social media companies to work with the Electoral Commission to strengthen their policies related to campaign advertising and material for elections and referendums, as well as to label the related ads to make their source clear and to ensure that their political ads databases are aligned with the country's regulations in this area.

BOX 37: *PUBLIELECTORAL*, ARGENTINA

The Argentinean NGO ***Asociación por los Derechos Civiles***, implemented an initiative called PubliElectoral, consisting of a digital tool that enables individuals to monitor electoral advertising in social media platforms. It consists in a plug-in that can be downloaded and installed on a computer's web browser. When navigating a Facebook page, the user can detect the political advertisements that they are being targeted with. The plug-in saves this information in a database that the Association uses as one of the inputs into its analyses assessing the transparency of online political advertising and related awareness-raising and advocacy. [420]

During the 2019 Argentinean elections, the organisation launched the first data collection of the PubliElectoral project. In it, they first sought to measure the variable «campaign deadlines». That is, if political parties complied with the deadlines related to electoral advertising according to the Argentinean electoral regime: if they started to advertise on networks once the primary and general election campaign had begun, and if they complied with the closed periods. For this purpose, the samples were taken in real time to be effectively representative of the periods. The stages of the project were designed according to the electoral calendar.

---

[419] See: https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/changing-electoral-law/transparent-digital-campaigning/report-digital-campaigning-increasing-transparency-voters.

[420] See: https://publielectoral.adc.org.ar/.

6.10. COOPERATION IN THE FIELD OF ELECTORAL CYBERSECURITY

As the frequency and impact of cybersecurity incidents increases and becomes even more dangerous for the integrity and safe conduct of elections, cybersecurity needs to be high up on the agenda of EMBs. This entails general risk management of these types of incidents, including prevention strategies and the development of worst-case scenarios. A lot of such preparation calls for inter-agency collaboration, including general cyber incident management (e.g., with Computer Emergency Response Teams, CERT) and professional media management.[421] International collaboration is crucial in this area, as the Internet does not start or stop at a country's borders.

BOX 38: INTERNATIONAL COOPERATION ON CYBERSECURITY AHEAD OF ELECTIONS

A good example of collaboration in the pre-electoral stage includes the work that was carried out ahead of the European Parliament elections. A Cooperation Group was established under Directive (EU) 2016/1148 of the European Parliament and of the Council, which identified cybersecurity of elections as a common challenge. This Cooperation Group comprises the competent authorities responsible for cybersecurity in the EU Member States, the European Commission, and the European Union Agency for Network and Information Security ('ENISA'). It mapped existing national cybersecurity efforts focused on network and information systems used for elections. After identifying risks linked to an insufficient level of cybersecurity that could potentially affect the 2019 European Parliament elections, it developed a Compendium on Cyber Security of Election Technology, which collected technical and organizational measures based on diverse experiences and good practices. The Compendium provides practical guidance for cybersecurity authorities and EMBs.[422]

Moreover, during the process surrounding the European Parliament elections, the EMBs of the 28 European Member States regularly exchanged information on cybersecurity and other matters of concern, through the European Cooperation Network on Elections.[423]

An example from another region was the cybersecurity symposium held in 2018 by the Organization of American States (OAS) to facilitate the sharing of good practices as well as to enhance coordination in this field in the Americas.[424]

---

[421] S. Van der Staak and P. Wolf, ***Cybersecurity in Elections: Models of Interagency Collaboration***, International IDEA, Stockholm, 2019. Available at: https://www.idea.int/sites/default/files/publications/cybersecurity-in-elections-models-of-interagency-collaboration.pdf.

[422] NIS Cooperation Group, ***Compendium on Cyber Security of Election Technology***, 2018. Available at: https://www.ria.ee/sites/default/files/content-editors/kuberturve/cyber_security_of_election_technology.pdf; Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2018, Securing free and fair European elections, Brussels, 12.9.2018 COM(2018) 637 final, https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-free-fair-elections-communication-637_en.pdf, p. 6 (also see Chapter 3, section 3.1.3).

[423] European cooperation network on elections. Available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/eu-citizenship/electoral-rights_en#electionsnetwork.

[424] Organization of American States OAS, 2018, Cyber Security Symposium, Washington. Available at: https://www.oas.org/en/sms/cicte/prog-cybersecurity.asp.

ACTIVITY VI

The following activity has the objective of determining the participant's level of knowledge on tools for electoral administration to upgrade their capacity to tackle disinformation and enhance cybersecurity during elections, the relevance of public agreements with Internet Service Providers and IT companies, how political parties and candidates can collaborate to prevent and counter disinformation, the crucial importance of collaborative fact-checking, myth-busting, trust, and credibility-enhancing initiatives; the phases of social media monitoring, strategies to counter hate speech, self-regulatory approach and hybrid co-regulatory approach to online content moderation and international cooperation in the field of cybersecurity.

Suggested guiding questions for a discussion:

I. Can you identify ICT applications and AI-powered tools for electoral administration for the different phases of the electoral cycle?
II. Can you explain why verification of online content during elections is crucial? Please provide some examples of factchecking initiatives.
III. Which are the main strategies that EMBs and electoral stakeholders can employ to counter hate speech all along the electoral cycle? What is the role of political parties and candidates?
IV. Why is social media monitoring an essential tool within EMBs' strategic and crisis planning? Can you describe the different phases of social media monitoring?
V. Please explain the differences between a self-regulatory approach to online content moderation and a hybrid co-regulatory approach.
VI. Why is international collaboration crucial in the field of electoral cybersecurity?

7. CONCLUSIONS

7.1. TENSIONS GUARANTEEING HUMAN RIGHTS AND FREEDOM OF EXPRESSION

1. The Internet, social media platforms, Artificial Intelligence (AI) and social messaging have dramatically changed how information is produced, communicated, distributed and consumed worldwide.
2. The impact of these new digital technologies affects fundamental human rights, international standards and norms related to political participation, freedom of expression, peaceful assembly, the right to privacy, and political participation, affecting persecuted minorities, women, youth, activists, journalists and media personnel, and the public.
3. Artificial Intelligence contributes to the automation of data analysis and has an extraordinary potential to improve and support upholding of the democratic values and processes and institutions, including elections.
4. Algorithms can help predict some human behaviours accurately. They can anticipate people's preferences about different topics, including civic and political participation, and they can accordingly influence users' choices by imposing different agendas and particular interests.
5. However, AI risk of being fed/underpinned by already biased data threatens to deepen social divides as the large-scale use of manipulative methods causes serious social damage, undermining human rights and democracies worldwide.
6. What was initially a generally positive assessment about the role of social media networks during elections has changed in recent years, giving way to concerns about the risks that the generally unregulated digital sphere poses to electoral integrity and democracies.
7. A further problem arises when adequate transparency and democratic controls are lacking. Companies, governments, and political groups, among other key players, can interfere with the electoral cycle by using technology companies' services to their advantage to achieve certain electoral results without citizens necessarily being aware of the operation.
8. The use of algorithms by social media platforms is inaccessible to most governments, as it is unregulated, or only partially so, and remains in the hands of private parties. The so-called resonance effect is created by the repetition of suggestions that are sufficiently customized to everyone. In this way, local trends are gradually reinforced, leading to the "filter bubble" or "echo chamber effect".
9. While social media platforms allow political contestants to reach out to their voters and engage them more directly in their campaigns, television still remains the most imporant source for political news in most countries. However, the platforms increasingly complement this, while also serving to enhance the opportunities for citizens to retrieve information that is important for their voting decisions, which is particularly impactful where freedom of expression and access to information is restricted.

7.2. ELECTIONS AND SOCIAL POLARIZATION

10. The Internet, social media platforms, AI and social messaging can contribute to social polarization, resulting in the formation of separate groups that no longer understand each other and find themselves increasingly in conflict with one another. In this way, personalized information could unintentionally collaborate to undermine social cohesion. Arbitrary blocking or filtering of online content are two forms of Internet censorship that can also take place all along the electoral cycle, altering democratic processes.
11. Their dynamics within the electoral cycle are complex, problematic, and full of tensions. Understanding these dynamics is fundamental for electoral bodies and practitioners to

safeguard the integrity and the credibility of electoral processes, as well as the role of the news media all along the electoral cycle.

12. The new information paradigm of the digital era includes content pollution on a global scale, a by-product of this new reality, and includes three main types of false and misleading content: disinformation, misinformation and mal-information.
13. Online threats, intimidation, harassment, trolling campaigns, and cyber-bullying lead to real-world targeting, harassment, violence, and murder, even to alleged genocide and ethnic cleansing, and pose very real off-line threats, with a disproportionate impact on women and vulnerable groups.

7.3. TRUST IN MEDIA AND JOURNALISM

14. Trust in media and journalism was weakening long before the advent of social media, and this trend was not separate from declining trust in institutions and intermediary bodies. Nevertheless, the massive volume and reach of disinformation and misinformation dressed up as news and distributed via social media has inflicted further damage to journalism and undermined democracies worldwide. As a result, citizens struggle to discern what is true and what is false, and what is in the grey zone between. This calls for finding a balance between protecting the integrity of the right to vote while also ensuring that freedom of expression is not hindered in the process.
15. The key tasks of the media in any democratic society – to inform the public about matters of interest to society; to act as public watchdogs exposing wrongdoing, such as corruption, and to provide a shared forum for public debate – take on added importance in the context of elections.
16. Disinformation, misinformation and mal-information undermine the right to vote, including the right for voters to participate in democratic discourse and electoral processes free from manipulation.

7.4. SAFETY OF JOURNALISTS

17. Attacks against journalists and media representatives, including bloggers and citizen journalists, and specific threats and violence targeting women journalists, have increased in the past decades, and impunity regarding these crimes prevails. The dangers currently faced by media have severe implications for elections, as digitally intensified patterns of threats and violence against journalists and other actors involved during the electoral cycle intensify during electoral periods.

7.5. THE ROLE OF THE STATES

18. Some States are deliberately tarnishing the reputations of human rights defenders, civil society groups, journalists, political actors, judges, EMBs, etc., by posting false information about them or orchestrating harassment campaigns. Others use digital surveillance tools to track down and target rights defenders and other people perceived as critics.
19. Some States, regional blocs, businesses, academics, and other key actors have developed ethical guidelines to overcome injustice and discrimination. But guidelines, codes of conduct and voluntary compliance are not always, by themselves, a robust enough response to the scale of the problem.
20. A worrying trend is some governments' use of laws formulated vaguely to justify excessive online censorship and the surveillance of journalists. Actors who contribute to public debate consequently have an increased risk to suffer from tracking, hacking, and doxing, fake domain attacks, phishing, online harassment, and DDoS attacks, among others, with disproportionate impact on women.

7.6. THE ROLE OF THE JUDICIARY

21. The judiciary is pivotal to the integrity of the electoral processes and an important actor all along the electoral cycle. The role of the judiciary is crucial too regarding upholding freedom of expression online, and access to information, especially in the campaign period. Also, given the fast development of the Internet, social media, social messaging and AI, their implications for human rights and the complexities they entail, it is important to build awareness and capacities among judiciary actors on these matters.

7.7. IMPACT ON EQUALITY AND WOMEN'S RIGHTS

22. The «digital divide» defines the gap between individuals who have access to and the requisite skills to use and benefit from modern information and communication technology, and those who don't.[425] Some parts of the world remain segregated from the Internet and other technologies, and their vast potential to improve life experiences, due to a lack of digital literacy skills, low education levels, and inadequate broadband infrastructure.
23. Countries of the Global South and countries in transition might be particularly vulnerable to poor Internet access or lack of secure Internet access, which further exposes them to cyber-attacks and online interference in elections.
24. In addition, AI technologies can have significant gendered implications, including, among others, gender-based exclusion, algorithmic bias and discrimination, the reinforcement of gender stereotypes, and the objectification of women.
25. There is overwhelming, global evidence of ICTs being used to perpetrate a broad range of violent acts against women during elections and in public life, especially acts that inflict fear and psychological harm. ICTs may be used directly as a tool of intimidation by threatening or inciting physical violence against women candidates, women journalists, voters, or representatives.
26. Violence against women in Elections (VAWIE) remains one of the most serious obstacles to the realization of women's political rights today. It can virtually disenfranchise women in elections, with effects on society that multiply from the resulting democratic deficit. VAWIE-Online is an umbrella term that captures a broad range of gender-specific abusive, harassing, degrading and violent discourse circulating on the Internet or mobile technology across a range of intensities, from sexist slurs to direct threats of physical harm.

7.8. CYBERSECURITY AND THE ELECTORAL CYCLE

27. With the expanded use and dependency on ICTs, the risks of interference in and manipulation of democratic electoral processes also grow. Cyber-fuelled attacks can undermine the legitimacy of elections and the mechanisms to protect them. Effective cybersecurity then plays a critical role in the EMBs' operational planning.
28. In this context, cybersecurity relates to protecting Internet-connected systems, networks, software, and data from unauthorized exploitation, including the security of offline election technologies, and protecting the integrity of the electoral process from disinformation and influence operations.
29. Responses to disinformation, misinformation and mal-information must also be considered by EMBs during the strategic planning processes of the electoral cycle. The electoral operational plan must foresee all the necessary resources to combat the misleading content, especially during election years.

---

[425] See: https://news.un.org/en/story/2021/04/1090712.

30. With the advances in technology in recent years, electoral management has increased in complexity in ways that often cannot be tackled without the help of ICTs. EMBs need to invest in qualified personnel, train new employees, reinforce the technical capabilities of temporary staff and allocate an adequate budget to keep hardware and software up to date.
31. EMBs have much to gain from upgrading their capacities to harness the use of technology in diverse ways. In times of pandemic and decreasing voter turnout, EMBs face ever-rising pressure – from policymakers and citizens alike – to offer public services online.

7.9. DIGITAL CAMPAIGNING

32. Digital campaigning allows candidates to reach new voters, which is positive for electoral participation. It can also make it easier and cheaper for campaigners to communicate with citizens, explain their policies and political views. On the other hand, new techniques for reaching voters– including micro-targeting, data mining, data harvesting, and the creation of psychometric profiles – could reduce confidence in the integrity of elections.
33. There is a need to enhance the transparency of online political advertising, including its source of financing, targeting methodology, and levels of funding, as well as the accountability of technology companies to national legislatures and other regulatory organs. A collaborative approach is necessary at the international level concerning regulatory principles.
34. Remedies involving restrictions on free speech and on political and electoral rights are controversial, as they may limit fundamental rights in a democratic society. Indeed, some human rights activists and international institutions have insisted that the best response to hate speech is more speech.
35. When it comes to the normative framework that applies to social media and AI in elections, the analysis should consider the international standards enshrined in national, regional, and international frameworks, rules and regulations, and the terms of service and community standards that guide social media companies' self-regulation.

7.10. THE RIGHT TO PRIVACY

36. Political micro-targeting fuelled by AI, driven by aggregated personal data that is not always collected in lawful ways and coupled with limited users' protection, gives increasing power to big technology companies and governments to track people's conduct, views and contacts online. In this context, tensions between freedom of expression and the protection of the rights of others – including reputation, privacy, data protection, and intellectual property rights – have also increased. Especially during electoral periods, citizens need to be able to assess whether certain news content is reliable or not, and to distinguish fact from opinion. This would underpin more informed choices about the news they consume and the content they share, comment on, or reutilize.
37. EMBs in different countries have responded to disinformation by collaborating with technology companies, supporting fact-checking initiatives, fostering voter's education, identifying disinformation threats, undertaking related strategic and crisis planning, coordinating with other state bodies, etc.

7.11. REGULATION OF ONLINE CONTENT DURING ELECTORAL PERIODS

38. The issue of the regulation of online content during electoral periods entails multiple complexities. The global nature of the Internet makes attempts to regulate it difficult and social media platforms are often protected from liability in many jurisdictions. They are often considered primarily aggregators or carriers of content produced by others – rather than

publishers – and therefore hold no editorial responsibility. As social media companies' impact on democracies has become more evident, the notion of their total exemption from liability is being contested.

39. Private companies make internal content regulation decisions that are often automated, and entail limited human review, factors that challenge traditional practices of norm formation and enforcement. More and more countries are trying to enforce certain restrictions during electoral processes considered necessary to ensure free and fair elections, compatible with Article 19 of the ICCPR if they are proportional and non-discriminatory.

40. The adequate regulation of political communications is crucial to ensure a just and equitable space for public dialogue and access to information during elections. Current regulatory frameworks tend to lag in this field. In many cases, they do not cover online political advertisements, and transparency regarding them is not guaranteed through standard reporting requirements. Another critical problem is the misuse of private information by political campaigns, parties, social media companies, and other commercial organizations.

41. There is a need to enhance the transparency of online political advertising, including its source of financing, targeting methodology, and levels of funding, as well as the accountability of technology companies to national legislatures and other regulatory organs.

7.12. VOTER EDUCATION AND MEDIA AND INFORMATION LITERACY

42. Voter education is key to the integrity of the electoral process. It covers the electoral period, the voter registration phase, and the candidate selection process for the political parties. In the current context, voter education and media and information literacy are closely interrelated. It is very relevant for voters to understand the dangers of digitally spreading disinformation and hate speech and have fact-checking basics tools and competencies.

43. Integrating media and information literacy topics into voter education programs will strengthen the integrity of the electoral process. In doing so, EMBs should collaborate with other State agencies, CSOs, different types of media, educational institutions (formal, informal and non-formal), social media platforms, and international organizations to raise awareness about these issues.

44. The development of media and information literacy should accompany any regulatory, self-regulatory or co-regulatory approach to online content. The proliferation of disinformation and misinformation depends in part on people's inability to distinguish between true and false, which is why a key part of the solution lies in critical thinking.

45. Especially during electoral periods, citizens need to be able to assess whether certain news content is reliable or not, and to distinguish fact from opinion. This would underpin more informed choices about the news they consume and the content they share, comment on, or reutilize.

46. The strengthening of critical skills should go beyond textual content, and also extend to building the electorate's awareness about manipulation strategies that appeal to their emotions. Media and information literacy can also help individuals to identify hate speech and to learn how they can contribute to counteracting it online.

7.13. FACT-CHECKING AND CONTENT VERIFICATION

47. As online disinformation becomes more widespread, professional journalistic standards and the values that traditional media represent – such as verification of content and publication in the public interest – are crucial.

48. In recent years, a myriad of collaborative fact-checking initiatives has emerged, led by media outlets and independent journalists, as well as involving other stakeholders. These efforts

debunk false information, or rate content by truthfulness, among many other approaches. Some are national and others are international.

7.14. COUNTERING HATE-SPEECH

49. EMBs have a critical part in curbing the fast spread of hate speech through social media, considering that it constitutes a central component of some electoral campaigns. While EMBs are primarily concerned with constitutional provisions and legislation governing electoral processes, political parties, and the media's role during elections, they are becoming increasingly aware of the legal and regulatory instruments tackling hate speech.
50. The implementation of innovative research that applies AI and other technological developments to identifying and countering hate speech should have a significant role to play in responding to the identified emerging challenges all along the electoral cycle.

7.15. SOCIAL MEDIA MONITORING

51. Social media monitoring is an essential tool within EMBs' strategic and crisis planning. It allows for the anticipation of disinformation threats and for preparing appropriate responses. It can also offer a solid basis for advocacy efforts, both in the short-term and long-term, by promoting improved regulation.
52. The exercise of social media monitoring during the electoral process can be a very helpful tool, but it can also present challenges. To start with, although it is possible to monitor certain elements related to electoral campaigns (e.g., hate speech content), exhaustive monitoring of social media networks is impossible due to the speed, reach, and volume of the content produced, along with the unlimited number of existing pages, accounts, and websites. The lack of access from data of crucial social platforms also restricts the extent of the analysis.
53. The multi-media nature of the content disseminated (text, video, audios, etc.) also makes social media monitoring particularly complex. Additionally, the closed nature of conversations taking place in private groups limits the scope of monitoring that it is possible to undertake. At the same time, the monitoring of closed groups presents serious concerns regarding privacy or the anonymity of users' accounts.

7.16. YOUTH ELECTORAL PARTICIPATION

54. Youth are a key audience of outreach actions, voter education, and media and information literacy initiatives ahead of elections. Initiatives targeting youth should harness social media and digital technology, enabling young women and men to play an active role in democratic processes, including in using and interacting with media and ICTs.
55. Social media platforms have extraordinary youth reach as millions of young people use them. They are a particularly relevant source of information for first-time voters. This massive reach can potentially integrate voting and election information into young people's social lives, thereby normalizing electoral participation and promoting a culture of political engagement.

Failure to take action might result in the further shrinking of civic space, decreased participation, enhanced discrimination, and a continuing risk of lethal consequences – in particular for underserved or vulnerable groups, such as women, minorities and migrants.

8. SUGGESTIONS FOR POSSIBLE ACTION

This chapter provides a series of suggestions for stakeholders when considering how to address the various challenges of holding elections in the digital era. These are based on existing good practice and available guidance. However, they are not meant to be prescriptive or exhaustive and due consideration should be given to the variance across countries and contexts in the regulation, legislation and management of elections.

TO ALL STAKEHOLDERS

1. Based on UNESCO's ROAM Principles, which advocate for a human-Rights based, Open, Accessible Internet governed by Multi-stakeholder participation (ROAM)[426], the UNESCO 26 principles to enhance the transparency of internet platform companies[427] and other principles for internet governance as debated globally through, among other forums, the World Summit on the Information Society (WSIS) follow-up and the annual Internet Governance Forum (IGF)[428], a multidimensional approach is recommended for addressing the challenges to free and fair elections posed by the development of the Internet, social media, and Artificial Intelligence (AI).
2. Governmental actors, civil society, the technical community, Internet intermediaries and other private sector actors, the media and academia, among other stakeholders, should be involved in developing efforts to enhance the understanding of, as well as responses to, the impact of online disinformation, propaganda and hate speech on human rights (particularly freedom of expression, privacy and the right to participate in public affairs), democracy, civic participation, and media development, among other issues.
3. Collaboration and good practice sharing should be fostered among all concerned stakeholder groups, both nationally and globally, focused on safeguarding the integrity of elections in digital times.
4. The implementation of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity should be supported, including through interventions specifically focused on ensuring the protection of journalists and other media personnel during electoral periods.

TO INTERNATIONAL AND REGIONAL ORGANIZATIONS

5. Assist Member States in ensuring that regulations and policies addressing disinformation, misinformation, and hate speech spread online and via private messaging apps align with international freedom of expression and privacy standards.
6. Facilitate dialogue and normative standard-setting regarding online political advertisement and related personal data extraction and use.
7. Provide technical support to the Member States to develop legal, regulatory, and policy frameworks that enhance transparency in this area aligned with international human rights standards.
8. Support interdisciplinary research and monitoring mechanisms to evaluate the impact of online disinformation on elections – following a gender-sensitive approach – and the effectiveness of responses implemented against it by different actors and their human rights implications.
9. Collaborate with governments, CSOs, educational institutions, media outlets, and social media companies to advance media and information literacy, focusing on reaching youth and marginalized or vulnerable population groups.

---

[426] See: https://www.unesco.org/en/articles/internet-universality-principles-and-roam-x-indicators-presented-5th-congress-polish.

[427] See: https://unesdoc.unesco.org/ark:/48223/pf000037723.

[428] See: https://www.intgovforum.org/multilingual/#mag.

10. Support journalists' capacity-building to cover elections and counter disinformation – including through the production of professional public interest journalism – as well as independent fact-checking, the use of AI to reinforce the work of media actors, and efforts to enhance journalists' safety during electoral processes, including that of women journalists.
11. Promote expanded and affordable access to the Internet to help bridge the digital divide and develop skills needed to take advantage of the opportunities that such access offers.
12. Contribute to advancing international consensus and compliance with the need to prevent arbitrary blocking of online content and Internet shutdowns, which have shown to be a critical challenge during elections.
13. Consider supporting innovative approaches to social media platforms' self-regulation, such as the proposal to establish Social Media Councils at the national, regional, or international level; as well as initiatives centred around Codes of Practice to be signed, respectively, by Internet intermediaries, political candidates, and media actors ahead of elections.
14. Support the strengthening of media and electoral regulators' capacities to tackle deliberate dissemination of hate speech, disinformation, and propaganda (including through some media outlets) through training and information sharing.

TO STATE ACTORS

TO POLICYMAKERS AND LEGISLATORS

15. Develop, strengthen, and implement a legal, regulatory, and policy framework that ensures respect for freedom of expression online and offline, privacy, and the right to participate in public affairs, in line with international standards.
16. Any limitation to the right of freedom of expression should be in accordance with the three-part test of legality, legitimate purpose, and necessity as outlined in Article 19 of the ICCPR. Any law aimed at addressing disinformation and hate speech should comply with these standards.
17. When designing new regulatory initiatives to address online harm, policymakers should follow established and recognized guidelines such as those developed by the Global Network Initiative.
18. Repeal criminal defamation laws, which are disproportionately restrictive, and favour instead civil laws with sanctions that are not as significant as to have a chilling effect on free expression, and are strictly proportional to the harm caused. Defamation offences should not apply to the expression of an opinion. Those accused of them should have the opportunity to prove the truth of their statements as well as benefit from other defences, such as showing that the publication of a statement on an issue of public concern was reasonable.
19. Laws giving special protection to public figures should also be revoked as, for the sake of open public debate, these actors should tolerate a higher degree of criticism than citizens in general.
20. Abstain from imposing Internet shutdowns and other unwarranted or disproportionate measures restricting the free flow of information, including the filtering or blocking of content. Such a drastic measure would only be permissible under particular and extraordinary circumstances authorized by official derogation on account of the livelihood of the entire country (as opposed to one political group in power) being at stake, for instance, because of a cyber-attack.
21. Foster an enabling environment for free, independent, pluralist and diverse media, and promote affordable access to the Internet for all, such access being increasingly inextricably linked and necessary to freedom of expression, but also to other rights such as the right to participation, freedom of association, of assembly, the right to health, among others. Advance education to develop digital skills that permit citizens to harness the possibilities provided by the Internet and other ICTs.
22. Ensure that adequate safeguards are in place to prevent undue interference with media's editorial independence, and refrain from any attempt to influence, censor or meddle in the

activities of media outlets and journalists. This is critical for them to provide the electorate with accurate and reliable coverage of elections, among other matters of public interest.

23. Protect the safety of journalists, media workers, citizen journalists and bloggers, including in relation to emerging digital and AI-assisted threats and attacks, as well as the anonymity of their sources. Implement tailored measures to address gender-specific threats targeting women politicians, candidates, EMB staff members, journalists and activists. Take resolute steps to address the prevailing impunity for crimes against journalists.
24. Ensure that not only state organs, but also businesses operating under their territorial jurisdiction, respect human rights, as per the Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework[429].
25. Refrain from delegating online content regulation to social media companies and other Internet intermediaries (be it through law or political pressure) beyond what is provided for by international standards.
26. Governments should consider the risks to freedom of expression involved in imposing social media companies with too little or too much liability for user-generated content, given that it can incentivise excessive content moderation and removal, including through automated mechanisms.
27. Encourage Internet intermediaries to develop self-regulatory mechanisms following a multi-stakeholder approach and in compliance with international human rights law and standards, as well as the adoption of a code of conduct by candidates and political figures ahead of elections, in order to prevent and counter disinformation and hate speech.
28. Contribute to and help advance ongoing debates related to the increasing calls for regulations and policies to further social media companies' duties of transparency and their increased accountability towards their users (in connection to algorithmic transparency, online advertising, policies governing content curation – including for detecting, demoting, removing, promoting and distributing content). This intersects with UNESCO's work to develop transparency standards that can assist Member States in promoting accountability while balancing interests in confidentiality.[430]
29. Strengthen regulation as required to ensure that political advertising is at least as transparent online as it is offline, in view of the need for enhanced accountability and data protection safeguards in connection to data-driven electoral campaigning on social media – given the challenges posed by the widespread use of micro-targeting, segmentation and profiling of potential voters. Efforts to reinforce regulations and measures in this area should respect international freedom of expression guarantees.
30. Ensure that the fundamental right to privacy is respected online. Personal data should not be used without consent, including for the purposes of political advertising and micro-targeting. Individuals should be informed in a transparent manner why they are seeing certain advertisements and who has paid for them. Thus, all online and offline advertisements should be publicly available and easily searchable, with detailed information stating who bought them, the source of the funds involved, how much was spent on them, who saw them, and the specific targeting parameters that were used. Comprehensive data protection laws must be implemented and enforced and any loopholes that can be exploited by political campaigns should be closed. It is also important to put in place measures to enforce detailed and timely reporting to electoral authorities on campaign financing and advertising.
31. Update electoral legislation in order to address other new challenges in the digital age, including, for instance, the fact that digital political campaigning takes place outside the strict electoral period.

---

[429] See: https://media.business-humanrights.org/media/documents/files/reports-and-materials/Ruggie-protect-respect-remedy-framework.pdf.

[430] See: https://en.unesco.org/news/unesco-initiates-global-dialogue-enhance-transparency-internet-companies-release-illustrative.

32. Provide civil society actors with access to the decision-making processes related to policies and laws concerning disinformation, misinformation and mal-information, and allow them to take a significant role in their implementation.
33. Support fact-checking and monitoring initiatives, researchers, and journalists' privacy – preserving access to data from Internet companies, and the sharing of good practices for tackling online disinformation and hate speech during the election cycle.
34. Promote media and information literacy, including in relation to content focused on elections, opportunities and risks related to digital technologies, and in consideration of media's contribution to democracy, among others. Special attention should be paid to young audiences, including through formal, informal, and non-formal education. These actions should be implemented in cooperation with EMBs, CSOs, educators, media and other actors that may help maximize their reach and impact.
35. Support the use of AI to combat disinformation, for instance through publicly financed, certified, and controlled social bots that can undertake automated analysis of online content. Ensure that the use of these tools is transparent and consistent with human rights obligations, including by putting in place the necessary safeguards so that they are not creating or exacerbating bias or discrimination, nor are dependent on any political party, candidate, or interest group.
36. Ensure that EMBs and data protection agencies are able to act independently, are properly resourced and possess sufficient know-how and enforcement powers – such as the ability to issue adequate fines.

TO ELECTION MANAGEMENT BODIES (EMBs)

37. Develop knowledge and gain access to social media monitoring tools to strengthen the own capacity to track, analyse, and anticipate disinformation attacks. Work closely with CSOs and research institutions that conduct social media monitoring to create an early warning system to identify disinformation narratives, in order to prepare adequate responses and communication strategies.
38. Secure voting equipment and other critical elections infrastructure, harnessing ICTs while also putting in place mechanisms to protect voter registration data, ensure verification and transparency and address cyber-security threats in connection to the electoral process.
39. Enhance communications and voter education strategies on information integrity, including by promoting media and information literacy to build resilience against disinformation. Toward this aim, work closely with civil society, education institutions and media actors, and in close coordination with other relevant state agencies.
40. Monitor campaign spending by political parties and candidates and take effective measures to ensure transparency and proper oversight of online campaign advertising, ensuring that social media platforms report on campaign spending (either following requirements set in law or through voluntary compliance measures). The disclosure of information about online ads should not be limited to those paid by contestants in elections, but should also apply to issue/third-party ads. The measures adopted by social media platforms in this area should ensure that political ads and other forms of sponsored content are clearly distinguishable and readily recognizable as paid-for communication.
41. Develop effective instances of collaboration with Internet intermediaries, fact-checking organizations, journalists, and researchers to counter disinformation and hate speech during elections.

TO DATA PROTECTION AGENCIES

42. Effectively monitor and supervise the implementation of data protection legislation, by exercising their investigative and corrective powers, as well as by timely handling complaints in cases where a breach of the law might have occurred.

43. Develop, provide, and enforce binding guidance to ensure the protection of personal data, safeguard the privacy of citizen's voting choices and defend public discourse from manipulation, in light of exploitation of data during political campaigning that risks undermining electoral integrity – for instance through profiling and AI-powered micro-targeting that does not comply with international human rights standards.
44. Contribute to cross-agency cooperation to ensure safe and secure elections, working closely with EMBs and other relevant actors to guarantee that personal information is used responsibly, and that individuals' privacy is respected – including in relation to voter registration, voter authentication, voting and results transmission, as well as in terms of overseeing political parties' use of data to organize their campaigns, design and target their messages and place online adverts.
45. Increase their capacities to manage ICT systems and tackle emerging digital challenges, while also enhancing their investigative ability to quickly follow up threats to voters' privacy and to the overall electoral process – for example by being sufficiently prepared to assess Internet platforms' collection of privacy-invasive information for the purpose of micro-targeting, as well as the storage of Internet user's political opinions.

TO THE JUDICIARY

46. Protect the judiciary's own freedom of expression within the limits laid down by international law, when its very own independence is at stake.
47. Provide an independent and impartial instance to which individuals can resort when their fundamental rights are violated, including their rights to freedom of expression and privacy, whether this occurs online or offline.
48. When considering laws and cases regarding measures to combat online disinformation and hate speech, help ensure that international human rights standards are fully upheld. To do so, it is important that judicial actors develop the necessary knowledge and capacities to understand the human rights implications of the Internet, social media and AI development, including in relation to free and fair elections, and adopting a gender-sensitive perspective.
49. Impartially resolve electoral disputes and guarantee the integrity of elections, including by overseeing that these are conducted in alignment with the related international standards, national laws, and regulations.

TO SECURITY FORCES

50. Contribute to upholding a peaceful environment for the conduct of elections in line with international human rights standards, acting impartially, without favouring any specific candidate or political party.
51. Protect the rights of freedom of expression, participation, association, and assembly during elections, including by ensuring journalists' safety as well as by avoiding arbitrary actions in relation to legislation that criminalizes disinformation. Toward this goal, human rights training should be provided to security forces members, and internal media and social media guidelines should be developed to avoid situations of public misperception, such as opinions being published by individual members that could be taken as official statements.

TO MEDIA REGULATORS

52. Ensure the enforcement of a regulatory framework that respects media's editorial independence and freedom, fosters citizens' access to a diversity of information sources as well as fairness in media coverage, in accordance with international human rights standards.
53. Monitor media coverage of elections and oversee that the applicable rules are respected, taking corrective action as needed.
54. If mandated to receive and manage complaints from candidates and political parties

regarding unfair or unlawful media coverage, handle them in an accessible, transparent and timely manner, offering a prompt and effective remedy and ensuring that any sanction – if applicable – is proportionate to the gravity of the offence.

TO POLITICAL PARTIES, CANDIDATES AND POLITICIANS

55. Consider the adoption of a Code of Conduct regarding the electoral process, with special emphasis on social media campaigning, to agree on basic rules and a possible multi-party mitigation board in case any violation is brought to its attention.
56. Refrain from using disinformation, misinformation and mal-information tactics in any messaging and campaign advertising, and avoid making, supporting or encouraging statements that cause these to spread.
57. Speak out to raise awareness about the risks of disinformation amplified by political actors, avoid its use to stigmatize and discredit journalists and media (for instance by portraying them as political opponents, liars or foes), and publicly condemn all threats and attacks against them.
58. Refrain from hate speech, including gender-based hate speech, threats etc. targeting women and minorities.

TO THE MEDIA

59. Provide the electorate with diverse, accurate, balanced, and unbiased information on candidates, political parties and their platforms, as well as regarding key debated issues and the electoral process itself, contributing to informed voting and to voter turnout.
60. Avoid broadcasting content based on unverified information, rumours and with an intention to cause a scandal or to spread propaganda. If a media outlet decides that the dissemination of such content is somehow important, even though it cannot be verified, it should broadcast it accompanied by a warning that explains that the information is not verified. Clearly distinguish facts from opinion-based comments, which should not be presented as news.
61. Support and abide by self-regulation mechanisms that either cover a whole industry sector (e.g., press councils) or are set within individual media outlets (e.g., ombudspersons or listeners' and viewers' editors, etc.).
62. Develop coverage to expose disinformation and propaganda, especially focused on gender-responsive coverage during elections, and on the human rights implications of these phenomena and the responses to them.
63. Consider adopting guidelines or codes of conduct specifically focused on reporting during electoral periods.
64. Take advantage of the use of trustworthy and unbiased AI tools for reporting, content production and distribution, coupled with the development and application of related guidelines and policies. While also reflecting on AI's implications for journalistic practice and the safety of journalists, such tools can help contribute to the strengthening of press freedom in face of the related emerging challenges.
65. Identify whether a news story has been fact-checked or not and implement fact-checking tools that clearly show whether this has been the case. Also strive to give greater visibility on social media platforms to verified and professionally edited news content, thus fostering trust in independent journalism.
66. Consider expanding efforts and investment of resources on investigating disinformation, as well as on fact-checking and myth-busting during electoral periods, including by creating or contributing to national and/or international collaborative initiatives in this area.
67. Put in place mechanisms to facilitate transparency regarding online advertising, including political ads.

68. Ensure that their staff is equipped to protect themselves against risks related to reporting on disinformation, such as threats, online harassment, and physical violence, and follow a gender-sensitive approach when preparing and responding to these challenges.
69. Develop capacities in and be particularly attentive to all forms of violence against women all along electoral cycles.

TO CIVIL SOCIETY ORGANIZATIONS, ELECTORAL OBSERVERS AND ACADEMIA

70. Take stock of international good practices in social media monitoring, including to adapt and develop AI tools and other technology-aided approaches to identify disinformation or hate speech, or analyse large datasets.
71. Team up with media outlets and journalists to monitor online content, detect disinformation campaigns and to establish rapid reaction mechanisms to limit the impact of hate speech, misinformation and disinformation.
72. Conduct and/or contribute to research focused on the human rights implications of AI (including implications that impact women or underserved populations in particular), monitor the actual impact of AI in this regard and publicly expose related abuses.
73. Lead and/or collaborate to strengthen media and information literacy.
74. Undertake research into the regulatory and monitoring frameworks for online campaign spending and current data protection legislation, to provide recommendations for further strengthening electoral integrity.
75. Increase information exchange to deepen knowledge about emerging good practices, ongoing projects, and lessons learned to strengthen mechanisms to monitor political/electoral campaigns online.

TO INTERNET INTERMEDIARIES

76. Contribute to developing, in cooperation with national authorities and other relevant stakeholders, multi-dimensional approaches to counter hate speech and disinformation (including disinformation in online advertising), particularly in connection to elections, toward the establishment of clear principles and guidelines in this regard. The approaches to tackle these issues could include, for example, requirements for the provision of additional information related to online political ads during electoral periods, limits placed on online political advertising ahead of elections, efforts to enhance the visibility of reliable information sources, and expanding content moderation and fact-checking resources in the lead-up to electoral polls.
77. Improve their capacity to detect and deffuse disinformation. Deploy AI tools to identify factually wrong content, inform users about it, create awareness of verified information and diversify political discourse through the infiltration of echo chambers and filter bubbles.
78. Support research and development focused on appropriate technological tools to counter disinformation that users may opt for voluntarily.
79. Collaborate on the establishment of self-regulation mechanisms set ahead of electoral processes with the aim of curtailing disinformation, hate speech and manipulation via online political advertising, among other problematic issues; as well as on setting industry-wide norms for tackling these challenges.
80. Use algorithms that are based on ethical benchmarks, and be consistent with global standard setting instruments that are being developed following a multi-stakeholder approach, such as UNESCO's Recommendation on the Ethics of AI and the ongoing work of the Council of Europe's Ad Hoc Committee on Artificial Intelligence towards a legal framework for the design, development and application of AI.
81. Facilitate researchers' and independent observers' access to algorithms, to ensure that they meet the requisite standards of ethics and transparent use.

82. Review advertising models to ensure that they do not adversely impact the diversity of opinions and ideas, and work closely with EMBs to specifically improve the own policies and practices related to political campaigning and advertising in the context of elections and the holding of referendums.
83. Establish open access political ad libraries featuring structured and transparent information on micro-targeting and digital platforms' advertising regulations, including data on those third parties that try to circumvent the rules, and not only on those who comply with them. This requires different solutions for different platforms, and should provide observers, researchers, media and interested electoral stakeholders with real-time information about online political ads.
84. Adopt clear, pre-determined policies governing actions to restrict third-party content (such as deletion or moderation) that go beyond what is required by law. Those policies should be based on objectively justifiable criteria rather than on ideological or political goals and should, where possible, be adopted following consultation with users.
85. Take effective measures to ensure that their users can both easily access and understand their terms of service and the policies and practices that they have in place for the above-mentioned type of actions, including detailed information about how they are enforced. Make available clear, concise, and easy to understand summaries of or explanatory guides regarding those policies and practices.
86. Respect minimum due process guarantees when taking content moderation and removal actions: promptly notify users when content that they created, uploaded or host may be subject to an action of this kind and explain the rationale behind the decision; give the user an opportunity to contest it, subject only to legal or reasonable practical constraints; scrutinize claims made under content moderation policies before acting, and apply the related measures consistently.
87. Consider the risks involved in over-relying on automation, notably in connection to freedom of expression, ensure that human review accompanies the use of AI tools for content moderation.
88. Take stock of the experience of urgently responding to disinformation disseminated during the COVID-19 pandemic, and apply the lessons learned to dealing with disinformation in electoral contexts.
89. Cooperate with national authorities, media actors, CSOs and educators in countering disinformation and hate speech, including through media and information literacy efforts.
90. Support fact-checking initiatives and independent journalism, as well as research to expand knowledge about disinformation and the effectiveness of responses to it (including their own). Regarding this last aspect, further facilitate access to their Application Programming Interface Access (API), beyond reserving it for commercial interactions only. To help advance sound research and oversight of electoral/political campaigns, provide better, more precise, and more coherent data to accredited election observers and researchers.

SOFTWARE, DATA MINING AND ADVERTISING COMPANIES

91. Respect human rights when undertaking operations, as per the Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. Of relevance to the matters addressed in this Guide is particularly the respect of the rights of freedom of expression, privacy, and political participation.
92. Provide reliable access through Application Programming Interface Access (API), make public announcements of changes to their API before they occur, and make available documentation that makes working with the API more feasible, as well as user support services that can receive reports about bugs and answer questions from users.
93. Advertising companies should refrain from placing ads on websites that help circulate inflammatory language, foster discrimination against minority groups and amplify hate speech.

9. ANNEX

9.1. GLOSSARY OF TERMS

Artificial Intelligence (AI)[431] in the context of this handbook refers to technology that enables the collection and processing of vast masses of data which facilitates actors to target the electorate on a large scale with tailor-made political messages, automated messages, and social bots, but also with fake accounts and disinformation advertising and other content, in order to influence the course of the election.[432]

Backdoor is a method, often secret, of bypassing normal authentication or encryption in an IT system.

Bot is a device or piece of software that can execute commands, reply to messages, or perform routine tasks, such as online searches, either automatically or with minimal human intervention. While some bot traffic can promote voter education, some use of bots can have a negative impact.[433]

Chatbot is an Artificial Intelligence (AI) application that can imitate a real conversation with a user in their natural language. Chatbots enable communication via text or audio on websites, messaging applications, mobile apps, or telephone.[434]

Computer security incident response team (CSIRT), often called a **computer emergency response team** (CERT) or computer emergency readiness team is an expert group that handles computer security incidents.

Cyber-attack is a digital attempt targeting the availability, confidentiality and integrity of data, systems or networks.

Cybersecurity is the protecting of Internet-connected systems, networks, software and data from unauthorized access and exploitation as well as the security of offline election technologies and protecting the integrity of the electoral process from disinformation and influence operations.[435]

Deepfake is fabricated media produced using Artificial Intelligence. By synthesizing different elements of existing video or audio files, AI enables relatively easy methods for creating 'new' content, in which individuals appear to speak words and perform actions, which are not based on reality.[436]

Defacement is an attack on a website that changes the visual appearance or content of the site or a webpage.

A **denial-of-service attack** (DoS) is a cyber-attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting the services of a host connected to the Internet. Denial of service is typically

---

[431] UNESCO, Recommendation on the Ethics of Artificial Intelligence. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[432] F. Zuiderveen Borgesius, J. Möller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodo and C. de Vreese, 2018, Online Political Microtargeting: Promises and Threats for Democracy, ***Utrecht Law Review***, 14(1); J. B. Bullock, 2019, Artificial intelligence, discretion, and bureaucracy, ***The American Review of Public Administration***, 49(7); T. Chen, L. Ran and X. Gao, 2019, ***AI innovation for advancing public service: The case of China's first Administrative Approval Bureau***, Paper presented at the Proceedings of the 20th Annual International Conference on Digital Government Research.

[433] See: https://www.dictionary.com/browse/bot, https://www.cloudflare.com/learning/bots/what-is-a-bot/.

[434] See: https://sendpulse.com/support/glossary/chatbot.

[435] S. Van der Staak and P. Wolf, ***Cybersecurity in Elections, Models of Interagency Collaboration***, p. 10, International IDEA, Stockholm, 2019. Available at: https://www.idea.int/sites/default/files/publications/cybersecurity-in-elections-models-of-interagency-collaboration.pdf.

[436] See: https://electionfactcheck.news/glossary/.

accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled.

Disinformation is false and misleading information that is potentially harmful to human rights.[437] The motivations underlying it could be to make a financial profit, to have foreign or domestic political influence, or simply to cause trouble.[438]

In a **distributed denial-of-service attack** (DDoS), the incoming traffic flooding the victim originates from many different sources. This effectively makes it impossible to stop the attack simply by blocking a single source.

Echo chambers are the (digital) environment in which a person encounters only beliefs or opinions that coincide with their own, so that their existing views are reinforced, and alternative ideas are not considered.

The **European Union Agency for Network and Information Security** (ENISA) is tasked with improving network and information security in the European Union.

The **European Parliament** (EP) is the directly elected parliamentary institution of the European Union.

Hypertext Transfer Protocol Secure (HTTPS) is a protocol for secure communication over a computer network, and is widely used on the Internet. In HTTPS, the communication protocol is encrypted using Transport Layer Security (TLS), or formerly, its predecessor, Secure Sockets Layer (SSL). The bidirectional encryption of communications between a client and server protects against eavesdropping and tampering of the communication.

Internet Service Providers are enterprises that provide subscribers with access to the Internet.

An **Internet Protocol** (IP) is the principal communications protocol in the Internet protocol suite. Its routing function enables the Internet to work and essentially establishes the Internet. An Internet Protocol (IP) address is a numerical label assigned to each device connected to a computer network that uses the Internet Protocol for communication.

Information technology (IT) is the use of computers, storage, networking and other physical devices, infrastructure and processes to create, process, store, secure and exchange all forms of electronic data. The commercial use of IT encompasses both computer technology and telephony.[439]

A **local area network** (LAN) is a computer network that interconnects computers within a limited area such as a residence, school, laboratory, university campus or office building. A **virtual LAN** (VLAN) is any communication layer that is partitioned and isolated in a computer network at the data flow layer.

Malicious software (malware) is any software intentionally designed to cause damage to a computer, server, or computer network.

Mal-information is accurate information that is shared with the intent to cause harm or to benefit the perpetrator, often by moving private information into the public sphere.

Manufactured amplifications refer to efforts by actors external to an Internet company which seek to artificially boost the visibility and reach of particular content by manipulating search engine results, promoting hashtags or links on social media or other means. These are distinct from organic amplification that impacts content rankings or recommendations, and which

---

[437] See C. Wardle and H. Derakhshan, 2017, p. 20. Available at: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[438] See: https://en.unesco.org/publications/balanceact.

[439] See: https://searchdatacenter.techtarget.com/definition/IT.

results from the fundamental algorithmic design of Internet services and follows corporate objectives.

Misinformation is false or misleading information that is shared without the intent to cause harm or the realization that it is incorrect. In some cases, actors may unknowingly perpetuate the spread of disinformation by sharing content they believe to be accurate among their networks.

Microtargeting can be defined as a "marketing strategy that uses people's data — about what they like, who they're connected to, what their demographics are, what they've purchased, and more — to segment them into small groups for content targeting."[440]

The Directive on Security of **Network and Information Systems** (NIS Directive) was set into policy by the European Parliament in 2016 in order to create an overall higher level of cybersecurity in the European Union.

Online platforms include a range of services available on the Internet including marketplaces, search engines, social media, creative content outlets, app stores, communications services, payment systems, etc.

Over-the-Top online services (OTTs) include the various content and communications related services provided by operators which piggyback on basic internet connectivity and access thereto.

Phishing is the fraudulent attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money), often for malicious reasons, by disguising as a trustworthy entity in an electronic communication.[441] **Spear phishing** is directed at specific individuals or companies, where attackers typically gather personal information about their target to increase their probability of success.

Security information and event management (SIEM) are software products and services that provide the real-time analysis of security alerts generated by applications and network hardware.

Security Operations Centre (SOC) is a facility that houses an information security team responsible for monitoring and analysing an organization's security posture on an ongoing basis. The SOC team's goal is to detect, analyse, and respond to cybersecurity incidents using a combination of technology solutions and a strong set of processes.[442]

Social bot is a technological programme that communicates more or less autonomously on social media, often with the task of influencing the course of discussion and/or the opinions of its readers.[443]

Social media are web or mobile-based platforms that allow for two-way interactions through user-generated content (UGC) and communication. Social media are therefore not media that originate only from one source or are broadcast from a static website. Rather, they are media on specific platforms designed to allow users to create ("generate") content and to interact with the information and its source.[444] Note that content on social media may be public or private, in regard to what is visible to other users of the service. Social messaging and closed social groups represent opaque social networks (although the platform owners may have a range of access to participants'

---

[440] See: https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/.

[441] A. J. Van der Merwe, M. Loock and M. Dabrowski, 2005, ***Characteristics and Responsibilities involved in a Phishing Attack***, Winter International Symposium on Information and Communication Technologies, Cape Town.

[442] See: https://digitalguardian.com/blog/what-security-operations-center-soc.

[443] E. Ferrara, O. Varol, C. Davis, F. Menczer and A. Flammini, July 2016, ***The Rise of Social Bots***, Communications of the ACM, 59 (7): 96.

[444] S. Kaiser, 2014, ***Social Media A Practical Guide for Electoral Management Bodies***, International Institute for Democracy and Electoral Assistance (IDEA), p. 11. Available at: https://www.idea.int/sites/default/files/publications/social-media-guide-for-electoral-management-bodies.pdf.

metadata and even the content data). Advertising and algorithmic feeds¸ recommendations and search results may also be rather opaque in that these may be unique to each individual user.

Strategic communication (STRATCOM) means organizational communication and image management that satisfies a long-term strategic goal of an organization or individual. **Trolls** are human-controlled accounts performing bot-like activities or harassing others online.

Violence against women in elections online (VAWIE-Online) is an umbrella term that captures a broad range of abusive, harassing, degrading and violent discourse circulating on the Internet or mobile technology across a range of intensities, from sexist slurs to direct threats of physical harm.[445]

A **virtual private network** (VPN) extends a private network across a public network and enables users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network. To ensure security, data travel through secure tunnels and VPN users use authentication methods – including passwords, tokens, and other unique identification methods – to gain access to the VPN.[446]

9.2. SELECTED INTERNATIONAL STANDARDS[447]

9.2.1.UNIVERSAL DECLARATION OF HUMAN RIGHTS

The Declaration was proclaimed by the United Nations General Assembly in Paris on 10 December 1948 (General Assembly resolution 217 A).

Article 19
Everyone has the right to freedom of opinion and expression; this right includes the freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

Article 20

1. Everyone has the right to freedom of peaceful assembly and association.
2. No one may be compelled to belong to an association.

Article 21

1. Everyone has the right to take part in the government of his country, directly or through freely chosen representatives.
2. Everyone has the right of equal access to public service in his country.
3. The will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures.

9.2.2. INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS

Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 entry into force on 23 March 1976, in accordance with Article 49.

Article 17

1. No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home

---

[445] ***Violence Against Women in Elections Online: A Social Media Analysis Tool***, IFES, 2019. Available at: https://www.ifes.org/sites/default/files/violence_against_women_in_elections_online_a_social_media_analysis_tool.pdf.

[446] A. G. Mason, 2002, Cisco Secure Virtual Private Network, Cisco Press, p. 7.

[447] The selected relevant articles are not replicated in full.

or correspondence, nor to unlawful attacks on his honour and reputation.

2. Everyone has the right to the protection of the law against such interference or attacks.

Article 19

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

   (a) For respect of the rights or reputations of others;

   (b) For the protection of national security or of public order, or of public health or morals.

Also see: Interpretative Document: General Comment 34 on Article 19 ICCPR Human Rights Committee 102nd session Geneva, 11-29 July 2011

Article 20

1. Any propaganda for war shall be prohibited by law.

2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.

Also see: Interpretative Document: General Comment No. 11: Prohibition of propaganda for war and inciting national, racial or religious hatred (Art. 20): 29/07/1983. CCPR General Comment No. 11.

Article 21

The right of peaceful assembly shall be recognized. No restrictions may be placed on the exercise of this right other than those imposed in conformity with the law and which are necessary in a democratic society in the interests of national security or public safety, public order (ordre public), the protection of public health or morals or the protection of the rights and freedoms of others.

Also see: Interpretative Document: General Comment No. 37 on Article 21 (Right of Peaceful Assembly) of the International Covenant on Civil and Political Rights, adopted on 23 July 2020.

Article 22

1. Everyone shall have the right to freedom of association with others, including the right to form and join trade unions for the protection of his interests.
2. No restrictions may be placed on the exercise of this right other than those which are prescribed by law and which are necessary in a democratic society in the interests of national security or public safety, public order (ordre public), the protection of public health or morals or the protection of the rights and freedoms of others. This article shall not prevent the imposition of lawful restrictions on members of the armed forces and of the police in their exercise of this right.
3. Nothing in this article shall authorize States Parties to the International Labour Organisation Convention of 1948 concerning Freedom of Association and Protection of the Right to Organize to take legislative measures which would prejudice, or to apply the law in such a manner as to prejudice, the guarantees provided for in that Convention.

Article 25

Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2 and without unreasonable restrictions:

(a)To take part in the conduct of public affairs, directly or through freely chosen representatives;

(b)To vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors;

(c)To have access, on general terms of equality, to public service in his country.

Also see: Interpretative document: General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25): 12/07/96. CCPR/C/21/Rev.1/Add.7, General Comment No. 25. (General Comments)

9.2.3. REGIONAL COMMITMENTS

African Charter on Human and Peoples' Rights: Date entry into force: October 21, 1986

Article 9

Every individual shall have the right to receive information.

Every individual shall have the right to express and disseminate his opinions within the law.

Article 10

Every individual shall have the right to free association provided that he abides by the law.

Subject to the obligation of solidarity provided for in Article 29, no one may be compelled to join an association.

Article 11

Every individual shall have the right to assemble freely with others. The exercise of this right shall be subject only to necessary restrictions provided for by law, in particular those enacted in the interest of national security, the safety, health, ethics and rights and freedoms of others.

Article 13

Every citizen shall have the right to participate freely in the government of his country, either directly or through freely chosen representatives in accordance with the provisions of the law.

Also see:

- Declaration of principles of freedom of expression in Africa
- SADC Principles and Guidelines Governing Democratic Elections
- American Convention on Human Rights (Adopted at the Inter-American Specialized
- Conference on Human Rights, San José, Costa Rica, 22 November 1969)

Article 11. Right to Privacy

1. Everyone has the right to have his honor respected and his dignity recognized.
2. No one may be the object of arbitrary or abusive interference with his private life, his family, his home, or his correspondence, or of unlawful attacks on his honor or reputation.
3. Everyone has the right to the protection of the law against such interference or attacks.

Article 13. Freedom of Thought and Expression

1. Everyone has the right to freedom of thought and expression. This right includes freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other medium of one's choice.
2. The exercise of the right provided for in the foregoing paragraph shall not be subject to prior censorship but shall be subject to subsequent imposition of liability, which shall be expressly established by law to the extent necessary to ensure:

a. respect for the rights or reputations of others; or

b. the protection of national security, public order, or public health or morals.

3. The right of expression may not be restricted by indirect methods or means, such as the abuse of government or private controls over newsprint, radio broadcasting frequencies, or equipment used in the dissemination of information, or by any other means tending to impede the communication and circulation of ideas and opinions.

4. Notwithstanding the provisions of paragraph 2 above, public entertainments may be subject by law to prior censorship for the sole purpose of regulating access to them for the moral protection of childhood and adolescence.

5. Any propaganda for war and any advocacy of national, racial, or religious hatred that constitute incitements to lawless violence or to any other similar action against any person or group of persons on any grounds including those of race, colour, religion, language, or national origin shall be considered as offenses punishable by law.

Article 15. Right of Assembly

The right of peaceful assembly, without arms, is recognized. No restrictions may be placed on the exercise of this right other than those imposed in conformity with the law and necessary in a democratic society in the interest of national security, public safety or public order, or to protect public health or morals or the rights or freedom of others.

Article 16. Freedom of Association

1. Everyone has the right to associate freely for ideological, religious, political, economic, labour, social, cultural, sports, or other purposes.
2. The exercise of this right shall be subject only to such restrictions established by law as may be necessary in a democratic society, in the interest of national security, public safety or public order, or to protect public health or morals or the rights and freedoms of others.
3. The provisions of this article do not bar the imposition of legal restrictions, including even deprivation of the exercise of the right of association, on members of the armed forces and the police.

Article 23. Right to Participate in Government

1. Every citizen shall enjoy the following rights and opportunities:

   a. to take part in the conduct of public affairs, directly or through freely chosen representatives;

   b. to vote and to be elected in genuine periodic elections, which shall be by universal and equal suffrage and by secret ballot that guarantees the free expression of the will of the voters; and

   c. to have access, under general conditions of equality, to the public service of his country.

2. The law may regulate the exercise of the rights and opportunities referred to in the preceding paragraph only on the basis of age, nationality, residence, language, education, civil and mental capacity, or sentencing by a competent court in criminal proceedings.

Association of South East Asian Nations Human Rights Declaration, 2012

21. Every person has the right to be free from arbitrary interference with his or her privacy, family, home or correspondence including personal data, or to attacks upon that person's honour and reputation. Every person has the right to the protection of the law against such interference or attacks.

[...]

23. Every person has the right to freedom of opinion and expression, including freedom to hold

opinions without interference and to seek, receive and impart information, whether orally, in writing or through any other medium of that person's choice.

24. Every person has the right to freedom of peaceful assembly.

25. (1) Every person who is a citizen of his or her country has the right to participate in the government of his or her country, either directly or indirectly through democratically elected representatives, in accordance with national law.

    (2) Every citizen has the right to vote in periodic and genuine elections, which should be by universal and equal suffrage and by secret ballot, guaranteeing the free expression of the will of the electors, in accordance with national law.

Also see:

- Inter-American Commission on Human Rights' Declaration of Principles on Freedom of Expression

Inter-American Democratic Charter (Adopted by the General Assembly at its special session held in Lima, Peru, on September 11, 2001)

Article 3

Essential elements of representative democracy include, inter alia, respect for human rights and fundamental freedoms, access to and the exercise of power in accordance with the rule of law, the holding of periodic, free, and fair elections based on secret balloting and universal suffrage as an expression of the sovereignty of the people, the pluralistic system of political parties and organizations, and the separation of powers and independence of the branches of government.

Article 4

Transparency in government activities, probity, responsible public administration on the part of governments, respect for social rights, and freedom of expression and of the press are essential components of the exercise of democracy.

Article 5

The strengthening of political parties and other political organizations is a priority for democracy. Special attention will be paid to the problems associated with the high cost of election campaigns and the establishment of a balanced and transparent system for their financing.

Article 23

Member states are responsible for organizing, conducting, and ensuring free and fair electoral processes.

Member states, in the exercise of their sovereignty, may request that the Organization of American States provide advisory services or assistance for strengthening and developing their electoral institutions and processes, including sending preliminary missions for that purpose.

European Convention on Human Rights and Protocols, Rome, 4.XI.1950

Article 8: Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 10: Freedom of expression

1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.
2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

Article 11 Freedom of assembly and association

1. Everyone has the right to freedom of peaceful assembly and to freedom of association with others, including the right to form and to join trade unions for the protection of his interests.
2. No restrictions shall be placed on the exercise of these rights other than such as are prescribed by law and are necessary in a democratic society in the interests of national security or public safety, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others. This Article shall not prevent the imposition of lawful restrictions on the exercise of these rights by members of the armed forces, of the police or of the administration of the State.

Protocol to the Convention for the Protection of Human Rights and Fundamental Freedoms Paris, 20.III.1952

Article 3 Right to free elections

The High Contracting Parties undertake to hold free elections at reasonable intervals by secret ballot, under conditions which will ensure the free expression of the opinion of the people in the choice of the legislature.

The Charter of Fundamental Rights of the European Union (proclaimed on 7 December 2000, entered into force on 1 December 2009)

Article 11 Freedom of expression and information

1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.
2. The freedom and pluralism of the media shall be respected.

ASEAN Human Rights Declaration (adopted on 18 November 2012)

Article 23

Every person has the right to freedom of opinion and expression, including freedom to hold opinions without interference and to seek, receive and impart information, whether orally, in writing or through any other medium of that person's choice.

Article 25

1. Every person who is a citizen of his or her country has the right to participate in the government of his or her country, either directly or indirectly through democratically elected representatives, `in accordance with national law.
2. Every citizen has the right to vote in periodic and genuine elections, which should be by universal and equal suffrage and by secret ballot, guaranteeing the free expression of the will of the electors, in accordance to national law.

10. SELECTED BIBLIOGRAPHY

10.1. UNITED NATIONS

- International Covenant on Civil and Political Rights adopted 16 December 1966, entered into force 23 March 1976, 999 UNTS 171 (ICCPR).
- UN GA resolution of 27 June 2016 on the Promotion, protection and enjoyment of human rights on the Internet, A/HRC/32/L.20.
- UN HRC Resolution 20.8 of 5 July 2012 and 26/13 of 26 June 2014 on the promotion and protection of human rights on the Internet.
- UN Resolution adopted by the Human Rights Council on 1 July 2016 32/13. The promotion, protection and enjoyment of human rights on the Internet A/HRC/RES/32/13 (2016).
- General Comment No. 34 on Article 19: Freedoms of opinion and expression, CCPR/C/GC/34, 12 September 2011.
- General Comment No. 37 on Article 21 - on the Right to freedom of peaceful assembly (CCPR/C/GC/37), adopted 23 July, 2020.
- UN Guiding Principles on Business and Human Rights. They were adopted by the UN Human Rights Council in 2011. Available at: https://www.business-humanrights.org/en/un-guiding-principles.
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, on hate speech online, 9 October 2019, A/74/486.
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye of 6 April 2018.
- UN Special Rapporteur on Freedom of Opinion and Expression, the Organisation for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organisation of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, «Joint declaration on freedom of expression and "fake news", disinformation and propaganda», 3 March 2017, http://www.osce.org/fom/302796.
- UN Women Inclusive electoral processes: A guide for electoral management bodies on promoting gender equality and women's participation (2015), https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2015/inclusive-electoral-processes-en.pdf?la=en&vs=633.
- UNESCO Addis Ababa Declaration World Press Freedom Day 2019 "Journalism and Elections in Times of Disinformation", UNESCO World Press Freedom Day International Conference, held in Addis Ababa, Ethiopia, 1-3 May 2019.
- UNESCO Journalists' Safety Indicators: https://en.unesco.org/themes/safety-of-journalists/journalists-safety-indicators.
- UNESCO universality principles: https://en.unesco.org/internet-universality-indicators/background.
- UNESCO: Countering Online Hate Speech, published on 2015, available at: https://unesdoc.unesco.org/ark:/48223/pf0000233231.
- UNESCO: Elections and Media in Digital Times, McGonagle, Tarlach, Bednarski, Maciek, Francese Coutinho, Mariana, Zimin, 2019, https://unesdoc.unesco.org/ark:/48223/pf0000371486.
- UNESCO: Intensified attacks, new defenses: developments in the fight to protect journalists and end impunity, 2019, https://unesdoc.unesco.org/ark:/48223/pf0000371343.
- UNESCO: ROAM-X indicators, 2020, https://en.unesco.org/internet-universality-indicators/roamx-indicators.
- UNESCO (2018). Improving the information ecosystem to protect the integrity of elections: Conclusions. Colloquium. 8 February 2018. Paris: UNESCO, https://en.unesco.org/sites/default/files/633_18_gni_intergrity_of_elections_final_report_web.pdf.
- UN (2019). ***Internet shutdowns during elections: considerations for UN engagement - UN Internal Document***. Retrieved from UN DPA EAD.

10.2. ASEAN

- AICHR Consultation on Freedom of Opinion and Expression in ASEAN (Article 23 of the ASEAN Human Rights Declaration) Nusa Dua, Bali, 10 December 2019 https://aichr.org/news/the-2019-aichr-consultation-on-freedom-of-opinion-and-expression-in-asean-article-23-of-the-asean-human-rights-declaration/.
- ASEAN Human Rights Declaration, adopted 18 November, 2012.

10.3. EUROPEAN UNION

- Charter Of Fundamental Rights Of The European Union, 2012/C 326/02.
- Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data.
- EC (Venice Commission). (2019). Joint Report of the Venice Commission and of the Directorate of Information Society and Action against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections. (p.3.). CDL-AD (2019)016.
- EU Action Plan against Disinformation, published on 5 December 2018, available at: https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf.
- Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (2018): Securing free and fair European elections; Brussels, 12.9.2018 COM(2018) 637 final, https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-free-fair-elections-communication-637_en.pdf.

10.4. EUROPEAN COURT OF JUSTICE CASE LAW

- ***Data Protection Commissioner v. Facebook Ireland Ltd and Maximillian Schrems*** Case C-311/18, 23 July, 2020.
- ***Eva Glawischnig-Piesczek v. Facebook Ireland Limited***, Case C18/18, 3 October 2019.
- ***Google v. Spain***, Case C-131/12, Judgement of the Court (Grand Chamber) of 13 May 2014.
- ***L'Oreal SA and others v. Ebay International A.G. and others***, Case C324/09, Judgment of the Court (Grand Chamber) of 12 July 2011.
- ***Maximilian Schrems v. Data Protection Commissioner*** Case C-362/14 of the Court of Justice of the European Union, of 6 October, 2015.

10.5. COUNCIL OF EUROPE

- Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) adopted 4 November 1950, entered into force 3 September 1953, amended by Protocol No 11, European Treaty Series No 155, entered into force on 1 November 1998, which replaced Protocols 2,3,4,5,8,9,10 and repealed Articles 25 and 46 of the Convention) (ECHR).
- Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) and Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, regarding supervisory authorities and trans-border data flows (ETS No. 181) as updated by Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data- Amending protocol to the Convention for the Protection of Individuals with Regard to the Processing of Personal Data, adopted by the Committee of Ministers at its 128th Session in Elsinore on 18 May 2018.
- Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, adopted by the Committee of Ministers on

7 March 2018 at the 1309th meeting of the Ministers' Deputies.
- Parliamentary Assembly of the Council of Europe (PACE) Report 15028, 'Democracy hacked? How to respond?', 8 January 2020.

10.6. EUROPEAN COURT OF HUMAN RIGHTS CASE LAW

- ***Ahmet Yildrim v. Turkey***, Application No.3111/10, Judgment of the European Court of Human Rights of 18 December, 2012.
- ***Animal Defenders International v. the United Kingdom*** App. no. 48876/08 (ECtHR, 22 April 2013).
- ***Big Brother Watch and Others v. the United Kingdom*** (applications nos. 58170/13, 62322/14 and 24960/15).
- ***Delfi AS v. Estonia*** (Application No.64569/09), Judgment of the European Court of Human Rights of 16 June 2015.
- ***Handyside v. United Kingdom Judgment EcHR*** (Application no. 5493/72), Judgment of 7 September 1976.

10.7. OSCE

- OSCE/ODIHR (2013): Handbook for the Observation of New Voting Technologies, Warsaw.
- Retrieved from: https://www.osce.org/odihr/elections/104939.
- OSCE/ODIHR (2021): Guidelines for Observation of Election Campaigns on Social Networks, Warsaw. Retrieved from: https://www.osce.org/Observing_elections_on_Social_Networks.

10.8. AFRICAN UNION

- African Union Convention on Cyber Security and Personal Data Protection (Malabo Convention), (2014) https://au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf.

10.9. INTER-AMERICAN COURT

- ***Velásquez Rodríguez v. Honduras***, Petition to the Court, Inter-Am. Comm'n H.R., Case No. 7920 (Apr. 24, 1986).

10.10. OTHERS

- (EU) Code of Conduct on Countering Illegal Hate Speech Online (2016), adopted by Twitter, Facebook, Microsoft and YouTube. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en#theeucodeofconduct.
- Organisation of the American States OAS (2018): Cyber Security Symposium, Washington, https://www.sites.oas.org/cyber/en/pages/default.aspx.

10.11. NATIONAL LEGISLATION / NATIONAL CASE LAW

- ***Carafano v. Metrosplash.com Inc,*** US Court of Appeals for the Ninth Circuit, No 02-55658, 13 August 2003.
- Constitution of the Republic of Greece, in force 1975, as amended 2008.
- ***Klayman v. Obama et.al***, US District Court for the District of Columbia, 6 December 2013.
- Network Enforcement Act of Germany, Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken of 1 September 2017 (Federal Law Gazette I, p. 3352 ff. In force from 1 October 2017).

- UK Government Online Harms White Paper, 12 February 2020, available at: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.
- United States Telecommunications Act of 1996, Pub.L.No 104-104, 110 Stat, 56 as amended by 47 U.S.C) § 230 (2000)- 47 U.S. Code § 230 - Protection for private blocking and screening of offensive material.

10.12. ACADEMIC ARTICLES / BOOKS / REPORTS / THESES

- Achler, M. (2020). New technologies and the right to freedom of peaceful assembly and association. European University Institute: embargoed thesis: https://cadmus.eui.eu/handle/1814/67031.
- Clouser, M., Krimmer, R., Nore, H., Schürmann, C. and Wolf, P. (2014). The use of open source technology in elections. Resources on electoral processes. ***International IDEA,*** Stockholm.
- Global Digital Report 2018, We are social & Hootsuite, available at: https://digitalreport.wearesocial.com/.
- Haque, Z. and Carroll, D. (2020). Assessing the Impact of Information and Communication Technologies on Electoral Integrity. ***Election Law Journal: Rules, Politics, and Policy***.
- Krimmer, R., & Volkamer, M. (2005). Bits or Paper? Comparing Remote Electronic Voting to Postal Voting. In ***EGOV (Workshops and Posters)*** (pp. 225-232).
- Krimmer, R., Duenas-Cid, D., & Krivonosova, I. (2021). New methodology for calculating cost-efficiency of different ways of voting: is internet voting cheaper?. ***Public Money & Management***, 41:1, 17-26, DOI: 10.1080/09540962.2020.1732027.
- Krimmer, R., Ehringfeld, A., & Traxl, M. (2010). Evaluierungsbericht: E-Voting bei den Hochschülerinnen-und Hochschülerschaftswahlen 2009. ***Bundesministerium für Wissenschaft und Forschung***, Wien.
- Krimmer, R., Hammerschmidt, G., Husted, T., Kleinaltenkamp, M., Mikhaylov, S. J., Raffer, C., & Schmidt, C. (2021). ***Good-Practice-Beispiele der Digitalisierung öffentlicher Verwaltung im Ausland***. (Forthcoming).
- Kyritsis, D. (2012). Constitutional Review In Representative Democracy, 32 Oxford J. Legal.
- Martin-Rozumilowicz, B. and Kužel, R., Social Media, Disinformation and Electoral Integrity, IFES working paper, August 2019, https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.
- McGonagle, T., Bednarski, M., Francese Coutinho, M., & Zimin, A. (2019). Elections and media in digital times. ***In Focus Series***.
- MEMO 98 Monitoring of posts by political parties on Facebook, European Parliament Elections 2019, pubished on 9 June 2019, available at: http://memo98.sk/uploads/content_galleries/source/memo/ep-elections-2019/fb-monitoring-ep-elections_shorter-version_fin.pdf.
- Rabitsch, Armin (2020). Regulation of Social Media in Elections. Charles University Prague.
- Reppel, L. and Shein, E., Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions, published on April 2019, available at: https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf.
- Selva, M. (2019). Reaching for the off switch: Internet shutdowns are growing as nations seek to control public access to information. ***Index on Censorship***, 48(3), 19-22.

Modern democracy requires free, transparent, and inclusive elections that are organized on a regular basis. In addition, it also needs a free, independent and pluralistic media landscape and the respect of the right to freedom of expression for all. Indeed, the fundamental right to freedom of expression, which includes press freedom and access to information, is critical to ensure a genuine political participation and respect of human rights in democratic societies.

The development of digital technologies has created new opportunities for communication between citizens, politicians, and political parties - with information related to elections made easily and speedily available to citizens. However, this new digital environment is also synonym of a rise in disinformation and misinformation which are often circulating unhinged via Internet networks, risking in some case to put democratic processes in danger.

In this context, all actors involved in electoral processes have an essential role to play. Electoral management bodies, electoral practitioners, political parties, the media, and civil society organizations need to understand the scope and impact of Artificial Intelligence and social media in the electoral cycle. They need to make the best use of them to strengthen democratic processes and concrete methodologies to identify who instigates and spreads disinformation, as well as strategies to hamper it.

The present handbook provides practical tools for a range of key electoral stakeholders in response to these pressing needs. It seeks to contribute to the achievement of Sustainable Development Goal (SDG) 16, which focuses on peace, justice, and strong democratic institutions. It also offers a comprehensive overview of international and regional standards and commitments related to the rights to freedom of expression, access to information, political participation, and privacy issues, which are vital when considering the impact of the Internet, social media, and Artificial Intelligence on elections. It also maps a series of good practices implemented by diverse stakeholders worldwide. Finally, this handbook also outlines suggestions for possible action by various electoral practitioners at the frontline – serving as a practical toolkit for them.

unesco

9 789231 005305