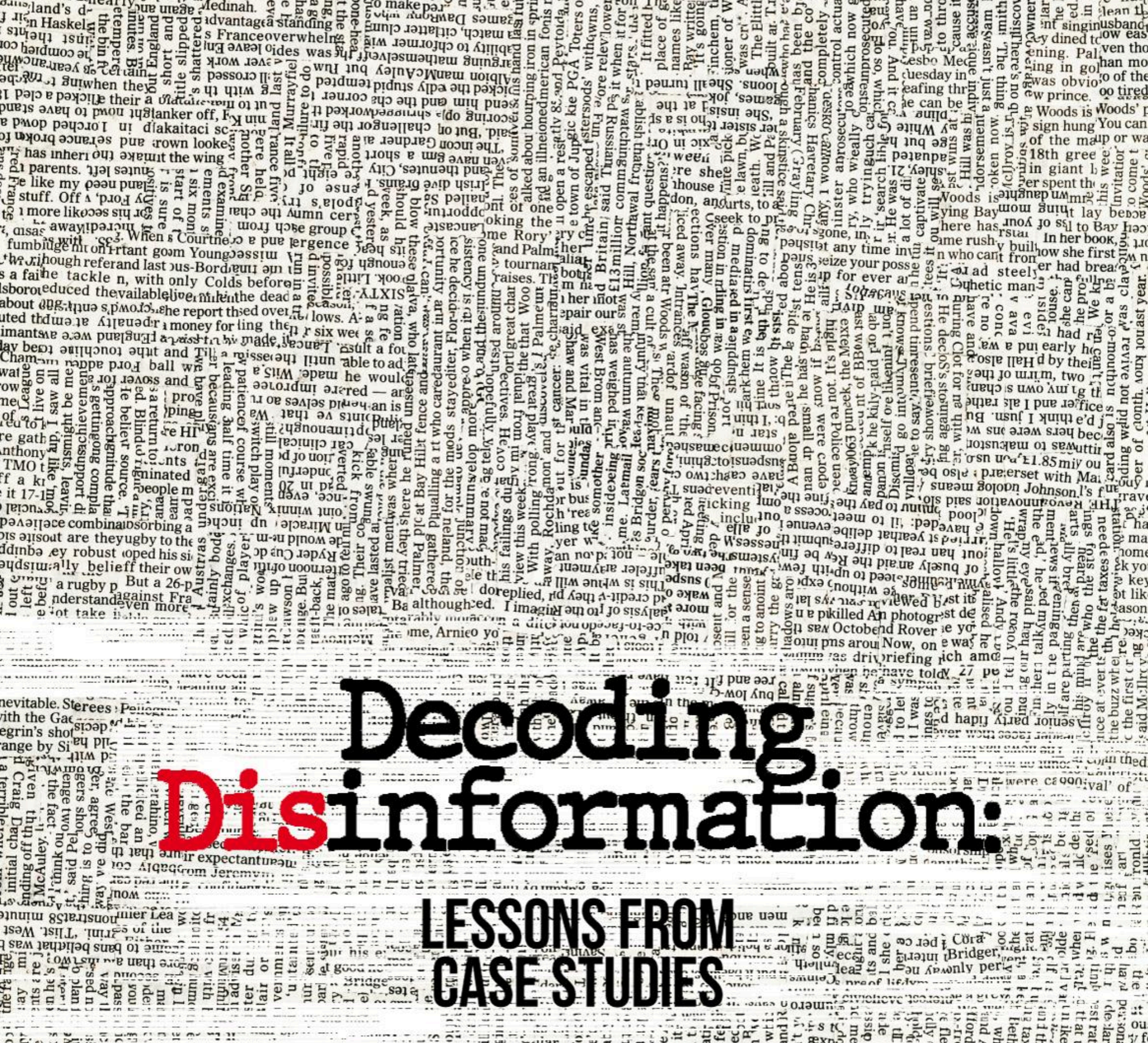


Decoding Disinformation:

LESSONS FROM CASE STUDIES

Authored by:
Boris Begovic, Bojana Kovac, Shita Laksmi and Marilia Maciel

cooperare germană
DEUTSCHE ZUSAMMENARBEIT

Implemented by
giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

DiPLO

Geneva Internet Platform

This report was developed with support of the project "Info Trust Alliance", funded by the German Federal Foreign Office and implemented by GIZ Moldova.

Decoding Disinformation: Lessons from Case Studies

Published by DiploFoundation/Geneva Internet Platform and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH Moldova

GIZ Office Moldova
73/1 Stefan cel Mare si Sfant Bvd. NBC - National Business Center, 9th floor
Chisinau, MD-2001
+373 22 996 031
giz-moldova@giz.de

DiploFoundation (Diplo)

Registered offices: Malta (headquarters), Geneva, Switzerland, Belgrade, Serbia and Washington, DC, the USA.

T+356 21 333 323
diplo@diplomacy.edu

Authored by: Boris Begovic, Bojana Kovac, Shita Laksmi and Marilia Maciel

Copy-editing: Nataša Grba Singh
Layout, design and illustrations: Diplo CreativeLab

Executive Summary

The notion of 'information disorder' encompasses a range of phenomena. Among them, mis- and disinformation stand out because of their potential to damage the public sphere, undermine democracy and negatively affect human rights. The complexity and scale of information pollution in the digitally connected world present an unprecedented challenge. In particular, social media has allowed information to be disseminated on a wider scale. While this new informational landscape has empowered individuals to express their opinions, it has also resulted in the spread of mis- and disinformation.

Although the level of exposure to disinformation on social media platforms is still a matter of debate, and more research is necessary, especially in countries from the Global South, studies have found that false information diffuses significantly farther, faster, and more widely than the truth. In addition, the emergence of artificial intelligence-generated mis- and disinformation introduces additional complexity. The challenges relate not only to misinformation fuelled by factual errors or fabricated information provided by AI but also to deliberate disinformation generated by malicious actors with the assistance of AI.

Against this background, a considerable number of national and regional legal frameworks, as well as private-led initiatives have been introduced to combat mis- and disinformation. On the one hand, they seek to empower individuals to participate in fighting the spread of mis- and disinformation through media literacy. On the other hand, there are initiatives that put in place content regulation aiming to protect society, with particular emphasis on vulnerable groups. In both cases, policies and frameworks to fight disinformation should seek to uphold human rights and fundamental freedoms.

This delicate balance lies at the heart of the present research, which investigates digital policy approaches that could help combat mis- and disinformation while protecting human rights. The research included case studies in four countries: Finland, Sweden, Lithuania, and Singapore. The takeaways could offer valuable insights to other countries and actors also seeking to curb the spread of mis- and disinformation. The key highlights from the research can be summarised in ten points:

1. **Disinformation is often described in broad, ill-defined terms**. This lack of clarity not only reduces the effectiveness of responses but also creates tensions between policies to combat information disorder and freedom of expression.
2. The issue of **mis- and disinformation is traversal and presents an interplay with several other digital policy areas**. Mapping these interplays is important in order to prevent potential unintended spillovers that policies to tackle false information may generate, as well as to identify pressure points that could be leveraged in the context of holistic policies to counter information disorder.
3. Any action to combat disinformation should be **aligned with international human rights law**, in order to protect the pillars of democratic societies.
4. Any advocacy of national, racial, or religious hatred that constitutes **incitement to discrimination, hostility, or violence is prohibited by international human rights law**, regardless of any assessment of its truthfulness. Public authorities and companies alike are under the **obligation to act** against such content.
5. Laws on disinformation that are **vague**, or that confer **excessive government discretion** to fight disinformation are concerning, and have led to censorship in some countries.
6. Some countries have achieved good results in combating disinformation without enacting specific domestic laws. Sweden, for example, has focused on the Psychological Defence Agency, an independent body with the resources to **monitor**

threats and map social vulnerabilities (such as growing discontent, which makes society particularly prone to fall prey to disinformation campaigns) and capabilities. The Agency also seeks to build long-term **societal resilience** against disinformation. Finland has achieved good results with an emphasis on **media literacy**, while Lithuania has relied on **online civic engagement** to debunk and prebunk disinformation.

7. Concerns with disinformation increasingly relate to **influence operations originating from abroad**. Dissociating external information influence campaigns from legitimate domestic opinion is difficult, especially in the context of **astroturfing** (the practice of hiding the sponsors of a message or organisation to make it appear as though it originates from, and is supported by, grassroots participants). Moreover, focusing on external operations may lead actors to **overlook genuine domestic societal vulnerabilities** that need to be addressed before they get maliciously exploited by domestic or foreign actors.
8. The introduction of **laws on disinformation should seek to protect legitimate and fundamental aims** – respecting the rights and reputations of others, protecting national security, public order, or public health or morals – and must be **legal, proportionate, and necessary**. Any limitation imposed on freedom of expression must be **exceptional and narrowly construed**.
9. More should be done to **curb economic incentives to disinformation**. Companies are expected to conduct human rights risk assessments and due diligence, ensuring their business models and operations do not negatively impact human rights. This includes **sharing data and information on algorithms**, which could make an assessment of the correlation between the spread of disinformation and ‘ad tech’ business models possible.
10. Companies should ensure that their **moderation practices are transparent**, consistent, and based on clear guidelines that **respect human rights**. A consistent and harmonised approach should also be fostered across platforms in order to avoid safe havens for disinformation.

Striking the right balance between protection and participation in combating disinformation means resorting wisely to both regulation and engagement. The latter should be conceived in broad terms, encompassing not only the active involvement of individuals, but also the involvement of other segments such as educators, companies, and technical actors. This inclusive approach provides a pathway to curb disinformation while respecting human rights.

Table of Contents

1.Introduction

Communication is the most basic element that holds the fabric of the social system together (Luhmann, 1992). Communicational units – composed of utterance, information, and understanding – serve not only as an element of self-reference, but also highly influence the communicational units to be uttered in response. By shaping the course of communication, agents influence the development of society.

The notion of 'information disorder' encompasses a range of phenomena (Wardle and Derakhshan, 2017). Among them, mis- and disinformation stand out because of their potential to damage the public sphere (Chambers, 2021), undermine democracy through amplification of distrust and polarisation (Hameleers, 2024) and negatively affect human rights (Colomina et al., 2021). Disinformation has also led to a decline in the credibility of factually accurate news (van der Meer et al., 2023), negatively affecting attitudes toward fact-finding and evidence-based research.

The deceptive use of information has a long history. Emblematic examples can be found in Ancient Egypt, during the Roman Empire, and after the invention of the printing press, for example. During the Cold War, the United States and the Soviet Union used disinformation campaigns to help advance their respective strategic interests (Ward et al., 2019). Cold War disinformation mitigation tactics led to the creation of government bodies and specific measures to counter disinformation. Although these initiatives may provide useful insights, they are not adequate to tackle the current wave of disinformation (Ward et al., 2019).

The complexity and scale of information pollution in the digitally connected world present an unprecedented challenge (Wardle and Derakhshan, 2017). The creation of the internet introduced significant changes in the informational landscape. It challenged the relevance of centralised information sources that used to be widely shared across societies, changing the production, distribution, and consumption of content. Information via digital media has become the crux of knowledge construction and identity formation (Frau-Meigs, 2024). In particular, social media has allowed information to be disseminated on a wider scale. While this new informational landscape has empowered individuals to express their opinions, it has also sometimes resulted in the spread of mis- and disinformation.

More research is necessary to understand the spread of disinformation and measure its impact on society, especially in countries from the Global South (Budak et al., 2024). The absence of sufficient data and research was also underscored by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (UNGA, 2021). So far, studies have found that false information diffuses significantly farther, faster, and more widely than the truth (Vosoughi et al., 2018; Glenski et al., 2018; Friggeri et al., 2014).

The speed of propagation is intimately related to the dynamics of social media. Individuals increasingly resort to social media for day-to-day information, but still use these platforms with a recreational mindset, which lowers critical thinking and makes them more vulnerable to content that provokes an emotional response, that has a powerful visual component or a strong narrative, or that is shown repeatedly (Wardle and Derakhshan, 2017).

Globally, data from 2022 shows that over 70% of individuals in some developing countries use social media as a source of news (Statista, 2022). This figure was above 60% in some European countries, such as Greece, Bulgaria, and Hungary. In the United States, 50% of adults get their news from social media (Khalid, 2019). In spite of that, their real level of exposure to disinformation on social media platforms is still a matter of debate. Budak et al. (2024), for example, argue that exposure is heavily concentrated among a small minority of people who already have extreme views and actively seek this type of content, while the average exposure to misinformation remains relatively low. The authors underscore that further research needs to be conducted in non-Western and non-English speaking countries to confirm these findings (Budak et al., 2024).

In 19 developed countries, 84% of Pew Research respondents believe that access to the internet and social media has made people easier to manipulate with false information and rumours (Pew Research, 2022). Moreover, 70% of those surveyed consider the spread of false information online to be a major threat, second only to climate change. As a reflection, Governments have used strong language to describe the problem of disinformation. For instance, the French Ministry of Culture has referred to disinformation as a 'threat to democracy' (Smith-Spark and Vandoorne, 2018).

In the United States, misinformation has also been labelled a threat to national security, especially following the 6 January attack, in which armed supporters of former President Trump stormed the US Capitol and killed five people (Healy, 2021). In 2019, India and Pakistan were on the brink of war after fake videos and pictures were disseminated across social media platforms (Phartiyal, 2019). The EU High Representative, Josep Borrell, remarked disinformation had become an industry and a potential weapon (EEAS, 2023). As the stakes get higher, so does the importance of more studies and empirical data to help governments and other stakeholders better assess the problem and identify the best strategies to overcome it.

The emergence of artificial intelligence-generated mis- and disinformation introduces additional complexity. The challenges relate not only to misinformation fuelled by factual errors or fabricated information provided by AI (often called AI 'hallucinations') but also to deliberate disinformation generated by malicious actors with the assistance of AI. The possibility to use generative AI models to produce 'deepfakes' - synthetic audio-visual media of human faces, bodies, or voices - enhances the quality and persuasiveness of disinformation, threatening core functions of democracy (Pawelec, 2022). Countries as diverse as Burkina Faso, India, Slovakia, Türkiye, and Venezuela have seen deepfakes used to sway voters and shape public opinion. Ultimately, deepfakes may undermine trust in elections and democratic institutions (Ray, 2021).

Against this background, a considerable number of national and regional legal frameworks, as well as private-led initiatives have been introduced to combat mis- and disinformation. On the one hand, they seek to empower individuals to participate in fighting the spread of mis- and disinformation through media literacy. On the other hand, there are initiatives that put in place content regulation aiming to protect society, with particular emphasis on vulnerable groups.

In both cases, policies and frameworks to fight disinformation should seek to uphold human rights, such as the right to freedom of expression, and the right to receive and impart information. The Human Rights Council has affirmed that responses to the spread of mis- and disinformation must be grounded in international human rights law, including the principles of lawfulness, legitimacy, necessity, and proportionality (UNGA, 2021).

This delicate balance lies at the heart of the present research, which investigates digital policy approaches that could help combat mis- and disinformation while protecting human rights and fundamental freedoms. This challenge requires multilevel, multidisciplinary, and multistakeholder efforts, as well as an examination of the full range of possible responses, which simultaneously leverage the power of laws, social norms, market forces, and technological architecture. This research unpacks the challenge by mapping the existing and emerging policy frameworks, identifying good practices, as well as strategic actors and partnerships.

This publication is composed of six sections. Section 1 is a general introduction. Section 2 presents the definition of dis- and misinformation adopted in the context of the present research. It also identifies the interplay between disinformation and other related digital policy areas such as infrastructure, security, economic issues, and human rights. Section 3 presents an overview of the two main approaches adopted by governments and other stakeholders to fight disinformation. The first (subsection 3.1) consists of promoting media literacy strategies that will foster the participation of individuals as frontliners in the fight against disinformation. The second (subsection 3.2) relates to the introduction of content policy regulation aimed at protecting society from threats posed by disinformation. Section 4 presents an overview of some initiatives to combat disinformation put in place specifically in the context of elections by governments (subsection 4.1) as well as by non-governmental actors (subsection 4.2).

Section 5 provides an analysis of the approach to combating disinformation adopted by four countries: Finland, Sweden, Lithuania, and Singapore. Section 6 presents some takeaways from the research. It focuses on the need for multidimensional and multistakeholder strategies to combat disinformation and presents some conclusions on the balance that must be achieved between tackling disinformation and upholding human rights and freedoms.

2. Problem definition and interplay with other areas of digital governance

Information integrity refers to the accuracy, consistency, and reliability of information (United Nations, 2023). Integrity is threatened by information disorder, a context in which a large variety of 'information pollution' is introduced in large volumes and with great velocity into civic discourse (Grabe and Bucy, 2023). The United Nations (2023) identifies three main vectors that threaten information integrity: disinformation, misinformation, and hate speech.

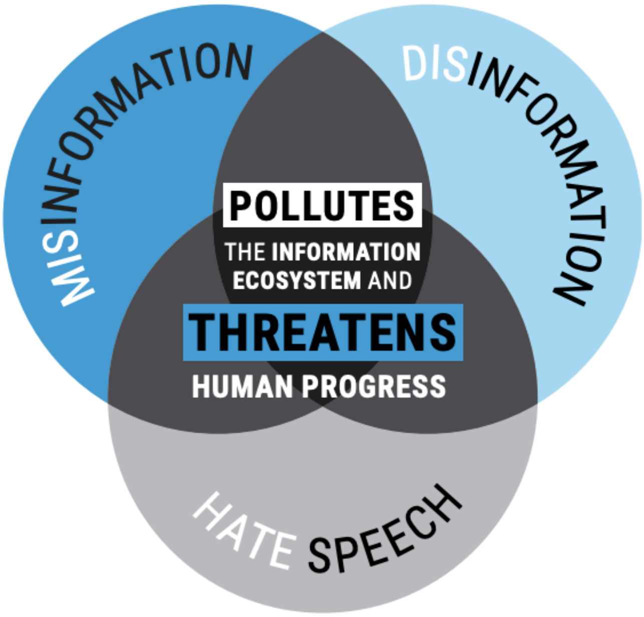


Figure 1. Threats to information integrity (adapted from the United Nations, 2023)

There is no consensus on how to define problems related to information disorder. The European Commission has described disinformation as verifiably false or misleading information that is cumulatively created, presented, and disseminated for economic gain or for intentionally deceiving the public, which may cause public harm (EU Commission, 2018a). The Broadband Commission (2020), on the other hand, has approached disinformation as false or misleading content with potential consequences, irrespective of the underlying intention or behaviours that produce and circulate messages. The way national laws and regulations approach the problem also varies.

According to the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, disinformation 'is often described in broad, ill-defined terms' (UNGA, 2021). This lack of clarity not only reduces the effectiveness of responses but also creates tensions between policies to combat information disorder and freedom of expression (UNGA, 2021). Laws designed to address vaguely defined concepts

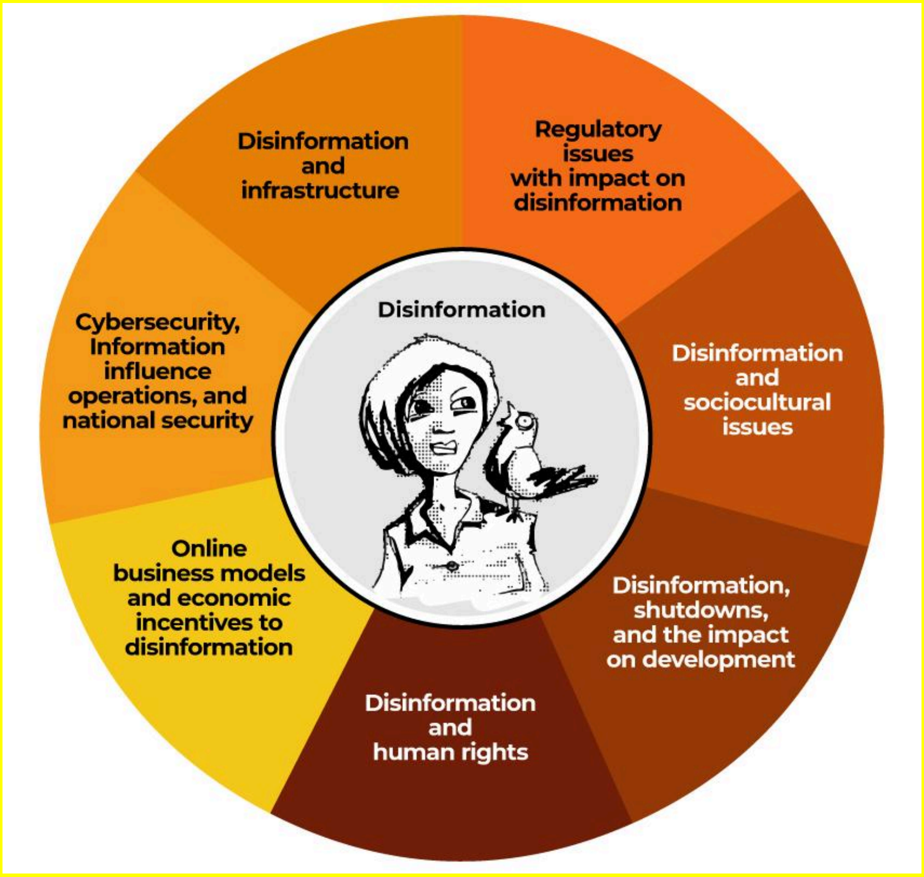
of ‘disinformation’ often contravene human rights law, lead to the criminalisation of permissible content, and significantly restrict information flows around the globe (UNGA, 2022, para. 60).
In the context of the present research, the following definitions will be employed:

- Misinformation is information that is false, but the person who is disseminating it believes that it is true (UNESCO, 2018).
- Disinformation is false information, and the person who is disseminating it knows it is false (UNESCO, 2018).
- Hate speech can be defined 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 on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor’ (United Nations, 2019).

This research focuses on mis- and disinformation, because of the deleterious impact they may have on the public sphere, blurring the distinction between facts and verifiable information on the one hand, and falsehoods and fabricated information on the other. They provoke a general state of mistrust and uncertainty within society that some authors have called the ‘post-truth’ predicament (Chambers, 2019; Braun, 2019).

The difference between mis- and disinformation lies with intent. While disinformation is intended to deceive and is spread with a hostile intent (to inflict harm), misinformation refers to the unintentional spread of inaccurate information shared in good faith. As pointed out by the United Nations (2023), however, the distinction between mis- and disinformation can be difficult to determine in practice. Misinformation can be rooted in disinformation, as deliberate lies and misleading narratives are weaponised, fed into the public discourse, and passed on unwittingly.

The issue of mis- and disinformation is traversal and presents an interplay with several other digital policy areas. It is possible to map these interplays by using the taxonomy of digital policy developed by Diplo and adopted by the Digital Watch Observatory of the Geneva Internet Platform. Mapping these interplays is important in order to prevent potential unintended spillovers that policies to tackle false information may generate, as well as to identify pressure points that could be leveraged in the context of holistic policies to counter information disorder.


Disinformation
Disinformation and infrastructure
Regulatory issues with impact on disinformation
Disinformation and sociocultural issues
Disinformation, shutdowns, and the impact on development
Disinformation and human rights
Online business models and economic incentives to disinformation
Cybersecurity, Information influence operations, and national security

2.1. Disinformation and infrastructure

In Diplo’s taxonomy, the infrastructure basket covers the technical aspects that form the backbone of the internet and the digital world. These include telecommunications and internet service providers (ISPs), critical internet resources (e.g. the domain name system and internet protocol numbers), and digital standards. While the interplay between these issues and mis- and disinformation is not immediately evident, there are a few indirect connections worth noting.

As a starting point, the online spread of mis- and disinformation would not be possible in the absence of the core internet infrastructure: just like all types of internet content, mis- and disinformation ‘travel’ online in the form of internet packets, and via different types of infrastructures (from Wi-Fi networks to submarine cables). The operators of these infrastructure elements often perform purely technical functions and typically have nothing to do directly with the content of the internet packets that travel via their networks.
In many jurisdictions, some of these operators (such as ISPs) are required by law to respect the net neutrality principle, treating all internet traffic equally, without prioritising or downgrading certain content, unless this is done for network management/optimisation related goals. In some cases, however, they may be called upon to take a hands-on approach and adopt content policy related measures for certain types of - most often - illegal content (e.g. blocking or throttling access to certain types of content). The impact that such a proactive approach would have on fighting disinformation will be discussed in the section dealing with internet intermediaries.

When it comes to digital standards, there is one particular area where they may have a direct impact on the spread or mis- and disinformation online. AI is increasingly being used to generate so-called synthetic content (in particular audio and video) that can be easily used to spread mis- and disinformation. In this context, technical standards are currently being developed to bring greater transparency in this area, by ensuring that users can easily identify whether the content they access is synthetic or generated with the help of AI. In June 2024, several standard-setting bodies – including, but not limited to the Content Authenticity Initiative (CAI), Coalition for Content Provenance and Authenticity (C2PA), IETF, IEC, ISO, and ITU agreed to establish 'a multistakeholder collaboration on global standards for AI watermarking, multimedia authenticity and deepfake detection technologies' (World Standards Cooperation, 2024).

Last, but not least, there have also been cases of mis- and disinformation related to internet/digital infrastructure issues, probably one of the best known cases being related to the so-called COVID-19 and 5G conspiracy theories' (Langguth et al, 2023), causally associating the COVID-19 outbreak with the introduction of 5G wireless technology.

2.2. Cybersecurity, information influence operations, and national security

States are increasingly turning to information warfare and information influence operations (IIOs) to achieve their strategic goals. State-sponsored disinformation can emanate from state institutions directly or from proxies, targeting audiences within the state's own territory or abroad for political and strategic aims (UNGA, 2021).

The interplay between disinformation and national security becomes particularly clear in the context of information campaigns originating from abroad, which may be defined as 'a set of activities coordinated by a foreign power that involves the promotion of misleading or inaccurate information or other specially-adapted actions aimed at influencing the decisions of politicians or other public decision-makers, the opinions of all or a part of the population, and opinions or decisions taken in other countries' (Swedish Psychological Defence Agency et al., 2024).

In influence campaigns, foreign countries study the controversies and challenges of a society, and exploit these vulnerabilities to polarise and disrupt, creating a climate of distrust. The threat posed by information influence lies in its potential to undermine critical democratic processes, and control public dialogue and decision-making. One of the most known examples of a disinformation campaign is the alleged Russian interference in the US elections and disinformation campaigns in 2016 and 2020 (Singh, 2020).

Disinformation particularly became the focus of cybersecurity discussions after the start of the war in Ukraine. On its digital frontlines were social media platforms, used to spread false narratives. One notable example is the first weaponised use of deepfake videos during an armed conflict – a fake video emerged on social media appearing to show Ukrainian president Volodymyr Zelensky asking Ukrainian troops to lay down their weapons. Another example includes the use of a deepfake of Putin declaring martial law, aired on Russian TV.

The right and duty of states to combat the dissemination of false and distorted news and the obligation of states to abstain from defamatory campaigns, vilification, or hostile propaganda – understood as interference in internal affairs – were officially brought up by UN Resolution A/RES/73/27 that established the UN open-ended working group on developments in the field of information and telecommunications in the context of international security. This call was introduced by Russia and its allies and states the importance of ensuring the credibility of information and combating fake news.

On the other hand, the USA and its allies believe that these issues fall under a different legal framework – that of the freedom of speech – and opt for distinguishing between the security of networks and policing content. They suggest that combating fake news should be addressed through public-private partnerships with the internet industry. Yet, the statements of some Western states (e.g. Germany, the Netherlands, and Australia) on the applicability of international law to cyberspace point out that disinformation campaigns of other states, which aim to alter election results and, thus, the political system of the state, may be considered a violation of state sovereignty.

At the same time, some countries are introducing combatting disinformation into their cyber policies; for example, the 2018 US National Cyber Strategy includes 'online malign influence and information campaigns and non-state propaganda and disinformation' (White House, 2018). Some countries are also establishing task forces and agencies that address disinformation. For example, the Swedish Psychological Defence Agency (PDA) was created to assist in detecting and resisting malign information influence directed at Sweden by antagonistic foreign powers.

2.3. Online business models and economic incentives to disinformation

Actors producing and distributing disinformation may have a range of motivations, such as financial, political, social, or psychological ones. In particular, financial motives and profiteering have been identified as significant incentives (Herasimenka et al., 2023). The monetisation of disinformation can be achieved in several ways. One of them is to seek direct funding from individuals, such as membership dues and donations to websites that spread false content. (Herasimenka et al., 2023).

In addition, there is a correlation between the economic incentives for disinformation and online advertising. According to the Global Disinformation Index (GDI), disinformation portals generate a sizable revenue from displaying online advertising on their own web pages. In Europe alone, disinformation sites are estimated to receive more than USD 76 million a year in revenues from online ads (GDI, 2020). Disinformation agents are further rewarded by the social media 'ad tech' industry.

Advertising is the main source of revenue for many tech companies, especially social media ones. This means that the maximisation of platform revenue depends on retaining users within the platform, consuming advertisements for the longest possible time. For this reason, platform algorithms are designed to prioritise eye-catching content, which may include sensationalist posts, clickbait, and disinformation.

While social media algorithms may play an important role in mainstreaming disinformation, there is little understanding of the importance of user preferences in this equation. Together, individual demand for content and algorithms shape the ads that will be displayed, but the weight of each of these elements is not entirely known. Some authors, such as Budak et al. (2024), argue that algorithms tend to push users to more moderate content and to offer extreme content predominantly to those who have sought it out. More transparency on the algorithms and the business of 'ad tech' would be necessary in order to assert the impact of social media on disinformation and propose adequate solutions.

2.4. Disinformation and human rights

The impact of disinformation on human rights is multifaceted, but it particularly affects the right to hold opinions and the right to freedom of expression. Article 19 of the Universal Declaration of Human Rights and of the International Covenant on Civil and Political Rights (ICCPR) protects the right to hold opinions without interference. This right is absolute and permits no exception or restriction. Information pollution leads to the blurring of the lines between facts and falsehoods, which restricts the capacity of individuals to freely form their own opinions. Disinformation, in particular, may lead to involuntary or non-consensual manipulation of the thinking process necessary to develop one's opinion.

The right to hold opinions and the right to freedom of expression are intertwined, since the former entails the capacity to freely access information necessary to form one's opinion and to change one's mind. Against this backdrop, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression affirms that state and non-state actors should not access and influence the thoughts and opinions of people without their knowledge or consent, such as restricting information and practising content curation through platform recommendations, tailored algorithms, or microtargeting (UNGA, 2021).

The right to freedom of expression is broad and encapsulates the freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers and through any media, offline or online. According to the UN Human Rights Committee (2011, para 47 and 49), the right to freedom of expression applies to all kinds of information and ideas, including those that may shock, offend, or disturb, irrespective of the truth or falsehood of the content. Freedom of expression may be restricted only in accordance with Article 19 (3) of the ICCPR, which requires all restrictions to be provided by law and to be necessary for the legitimate aim of respecting the rights and reputations of others, and for protecting national security, public order, or public health or morals.

Although disinformation has the potential to negatively impact other legitimate public objectives, as noticed during the COVID-19 pandemic, any limitation of disinformation must establish a close and concrete connection to the protection of one of the aforementioned legitimate aims (Art. 19 (3)). In addition, all restrictions must be exceptional and narrowly construed (UNGA, 2021). This means that, under international human rights law, fighting false information is not, in itself, a legitimate aim that justifies restricting freedom of expression. According to the Special Rapporteur, vague laws that confer excessive discretion to fight disinformation are particularly concerning, as they can lead to arbitrary decision-making (UNGA, 2021).

Disinformation may also affect other rights, especially during specific political processes, such as elections. Disinformation can distort the public perception and influence electoral outcomes, undermining the right to free and fair elections—a cornerstone of democratic governance.

At the same time, policies initially aimed at fighting disinformation may be easily misused and abused by public authorities allowing governments greater control and discretion over the public discourse, imposing arbitrary or politically motivated limits to freedom of expression (APC, 2021). Given the fundamental importance of freedom of expression to democracy and the enjoyment of all other human rights and freedoms, international human rights law affords particularly strong protection to expressions on matters of public interest, including criticism of governments and political leaders and speech by politicians and other public figures.

There is also a grey area between disinformation and hate speech, where disinformation may incite violence, fuel discrimination, and marginalise vulnerable groups. Disinformation campaigns often target ethnic minorities, immigrants, and other marginalised communities, exacerbating social tensions and leading to hate crimes. Article 20 (2) of the International Covenant on Civil and Political Rights provides that any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence is to be prohibited by law, regardless of any assessment of its truthfulness.

2.5. Disinformation, shutdowns, and the impact on development

The connection between unfettered access to the internet and development is well established (Hjort and Tian, 2024). Internet shutdowns are one of the issues that hinder free, stable, and reliable access to the internet, and bring about significant economic losses to businesses.

Internet shutdowns have emerged as an extreme, yet recurrent, practice to control online communication, including the spread of mis- and disinformation. In Africa, for example, both autocratic and democratic governments have increasingly resorted to shutdowns (CIPESA, 2019) in response to concerns about disinformation around elections, or when confronted with the potential for online hate speech to encourage violence (Gagliardone and Stremlau, 2022). This is detrimental not only to freedom of speech, but also to the development of the digital economy as can be seen in Figure 2.

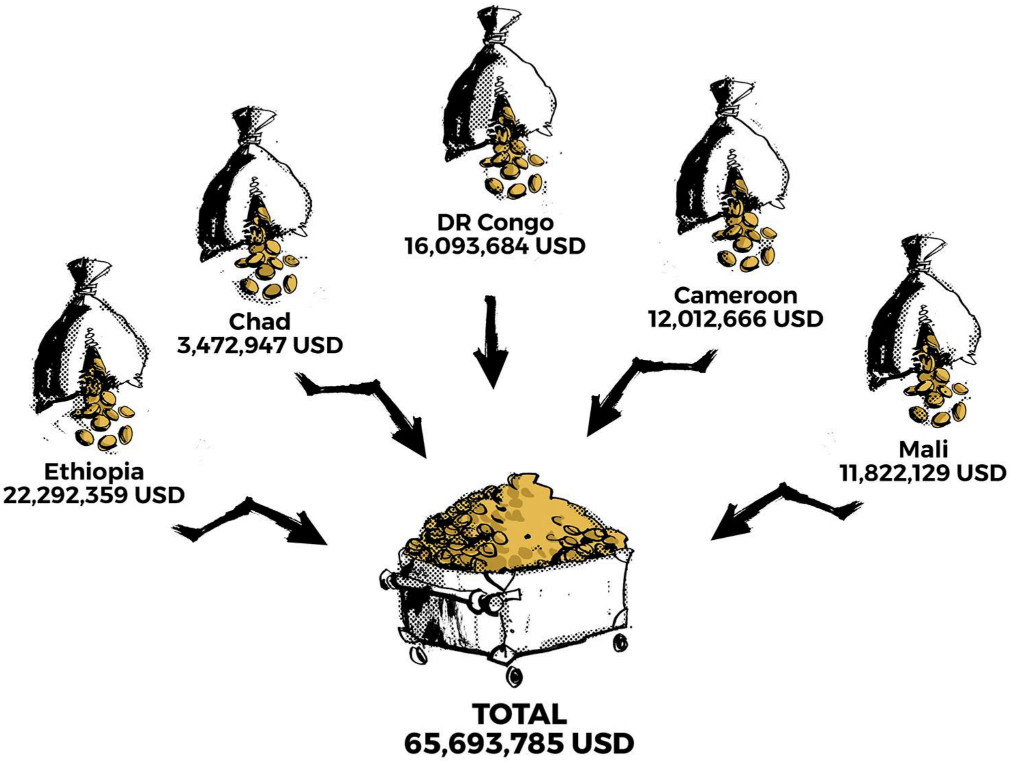


Figure 2. Estimated economic loss (USD) of a 5-day total shutdown in some African countries (adapted from CIPESA, 2019).

Internet disruptions, however short-lived, affect many facets of the national economy and tend to persist far beyond the period in which access is disrupted, undermining investor confidence, raising reputational risk, and harming foreign direct investments.

Measures to fight disinformation must be legal, proportionate, and necessary. Moreover, investing in human development through education and media literacy is important to ensure that citizens become active participants and frontliners in combating disinformation, weakening the urge to resort to shutdowns.

2.6. Disinformation and sociocultural issues

In Diplo's taxonomy, policy issues triggered by the impact of the internet on social and cultural life include a wide range of issues, from content policy to the promotion of cultural diversity and multilingualism. In the context of fighting disinformation, content moderation refers to policies oriented towards preventing the production and dissemination of disinformation, the removal of existing disinformation, or the downgrading of disinformation-related content through algorithms, so it becomes less visible to users.

Due to the significant increase in the spread of mis- and disinformation, a considerable number of national and regional legal frameworks, as well as private-led initiatives have been introduced to curb or remove false content. Although content policy has a key role in fighting disinformation, these measures should be considered in tandem with other areas, such as human rights protection and media literacy policy.

2.7. Regulatory issues with impact on disinformation

Legal issues are present in the interplay between fighting disinformation and any other digital policy area. For example, content policy largely depends on laws that determine the extent to which intermediaries, such as social media platforms, can be held liable for content uploaded by third parties.

Countries and regions have adopted different approaches to liability. While some jurisdictions have considered protecting intermediaries from liability as key to ensuring freedom of expression and innovation, others have adopted more stringent approaches, defining thresholds that will trigger liability, as well as measures that should be taken by intermediaries when illicit or harmful content is uploaded on their platforms. These different approaches to liability are also related to different understandings of the limits of freedom of expression across jurisdictions.

Data protection is also important in discussions about disinformation since data collection allows platforms to understand the interests and preferences of individuals. Targeted advertising based on profiling is a mechanism used to enhance the efficiency of disinformation campaigns. This means that high data protection standards can also collaterally contribute to the fight against mis- and disinformation.

The multidimensional nature of information disorder reveals the interplay between disinformation and several other digital policy areas. It also shows that a wide range of actors - governments, companies, civil society, educators - need to be involved in discussions about disinformation in order for holistic and effective policies to be conceived.

3. Policies and regulatory frameworks to combat mis- and disinformation

Policies introduced to combat mis- and disinformation are related to two main sets of goals: 1. promoting individual and societal resilience by strengthening agency and promoting participation, and 2. protecting society by reducing societal disinformation exposure. This two-pronged approach has led to more public engagement in shaping media literacy initiatives on the one hand, and more government regulation on the other (Frau-Meigs, 2024).

Policies that foster media literacy as a way of combating disinformation are underpinned by a participatory approach, in which the solution to the problem is necessarily co-constructed with the interpreters of the message. These policies aim to strengthen analytical skills and critical thinking, empowering individuals, and placing them as front-liners capable of ‘prebunking’ and ‘debunking’ practices (Jones-Jang et al., 2019; Hameleers, 2024).

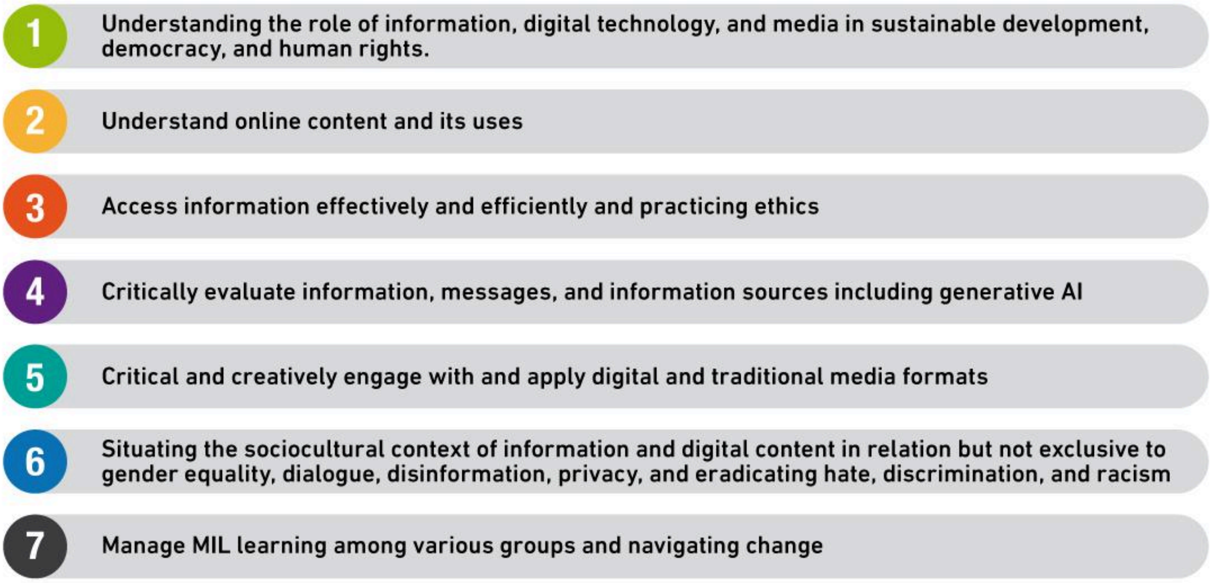
In parallel, regulation embodies a protectionist approach, which recognises the asymmetry of power between media systems and individuals. It aims to act upon the information environment, by reducing societal exposure to disinformation. Regulation often introduces certain obligations to media owners, or to protect certain segments of the population considered more vulnerable (Frau-Meigs, 2024). Children may be particularly susceptible to mis- and disinformation due to their evolving cognitive capacities (UNICEF, 2021a). If children are exposed to false information, there is a chance they may spread it without being aware of either the content or the consequences, underscoring the complementarity between regulation and participatory media literacy strategies.

3.1. Media literacy and education

Spotting mis- and disinformation online has become a major challenge. While technology increases the capacity of individuals to receive and impart information, it may also be misused to propagate false content faster, and to create it in a way that makes false information appear highly authentic. In this scenario, media and information literacy (MIL) has been identified as a key strategy to fight mis- and disinformation.

Traditional media literacy education is aimed at fostering individual awareness and strengthening critical thinking in relation to a broader set of media channels, such as newspapers and television (Frau-Meigs, 2024). In recent years, the notion of media and information literacy has put a sharper focus on the digital communication environment and on the competencies required to navigate it (Frau-Meigs, 2024).

UNESCO (2013) defined MIL as ‘a set of competencies that empowers citizens to access, retrieve, evaluate and use, create as well as share information and media content in all formats, using various tools, in a critical, ethical and effective way, in order to participate and engage in personal, professional and societal activities’. A list of seven non-exhaustive competencies related to media literacy can be found in Figure 3.



1. Understanding the role of information, digital technology, and media in sustainable development, democracy, and human rights.
2. Understand online content and its uses
3. Access information effectively and efficiently and practicing ethics
4. Critically evaluate information, messages, and information sources including generative AI
5. Critical and creatively engage with and apply digital and traditional media formats
6. Situating the sociocultural context of information and digital content in relation but not exclusive to gender equality, dialogue, disinformation, privacy, and eradicating hate, discrimination, and racism
7. Manage MIL learning among various groups and navigating change

Figure 3. Broad non-exhaustive media and information literacy competencies (adapted from Grizzle et al., 2021)

By promoting agency, media literacy fosters prebunking and debunking practices. While debunking involves exposing an already disseminated false claim, prebunking tackles disinformation before it has been spread. Audiences are ‘inoculated’ against misleading information, enabling them to recognise and prevent their amplification. The goal is to create ‘mental antibodies’ by exposing individuals to weakened versions of ‘fake news’ and strengthening their capacity to identify and resist this type of information (Roozenbeek and van der Linden, 2019).

This can be achieved, for example, by revealing the main mechanisms and techniques employed in disinformation strategies, and by using pedagogical tools, such as simulations and games. An example is Bad News, a free online game in which players take the perspective of a fake news tycoon. Media literacy is also important to debunking initiatives organised around fact-checkers that offer short, argument-based refutations of falsehoods.

Although there is initial evidence that a combination of prebunking and debunking strategies produces positive results in the fight against disinformation, more research is needed to measure the effectiveness of these strategies and their long-term effect, helping to better integrate them into media literacy policies not only at school, but also during lifelong education (Hameleers, 2024).

In focus: AI literacy in the context of MIL strategies

The growing use of artificial intelligence (AI) creates further challenges in the fight against mis- and disinformation. The ease with which AI can generate and spread false information surpasses the capabilities of traditional regulatory and oversight mechanisms. AI literacy can help individuals discern truth from falsehood. In the context of fighting mis- and disinformation AI literacy is structured in three pillars:

- **Enhancing critical thinking**. It is important to raise awareness about the capabilities of AI to produce convincing, yet false, information. This involves teaching individuals to critically assess online information, distinguish between reliable and unreliable sources, and recognise the mechanisms through which AI can generate and propagate false information.

- **Demystifying the technology itself.** This includes providing basic knowledge of how AI algorithms are trained and operate, providing individuals with the insights needed to question AI-generated content.
- **Instilling a sense of ethical responsibility.** Integrating discussions about AI's ethical implications, including its role in misinformation, helps students consider the broader societal impacts of AI. This entails questions about who is accountable when AI is used to misinform, and the ethical obligations of AI developers to prevent misuse of their technologies.

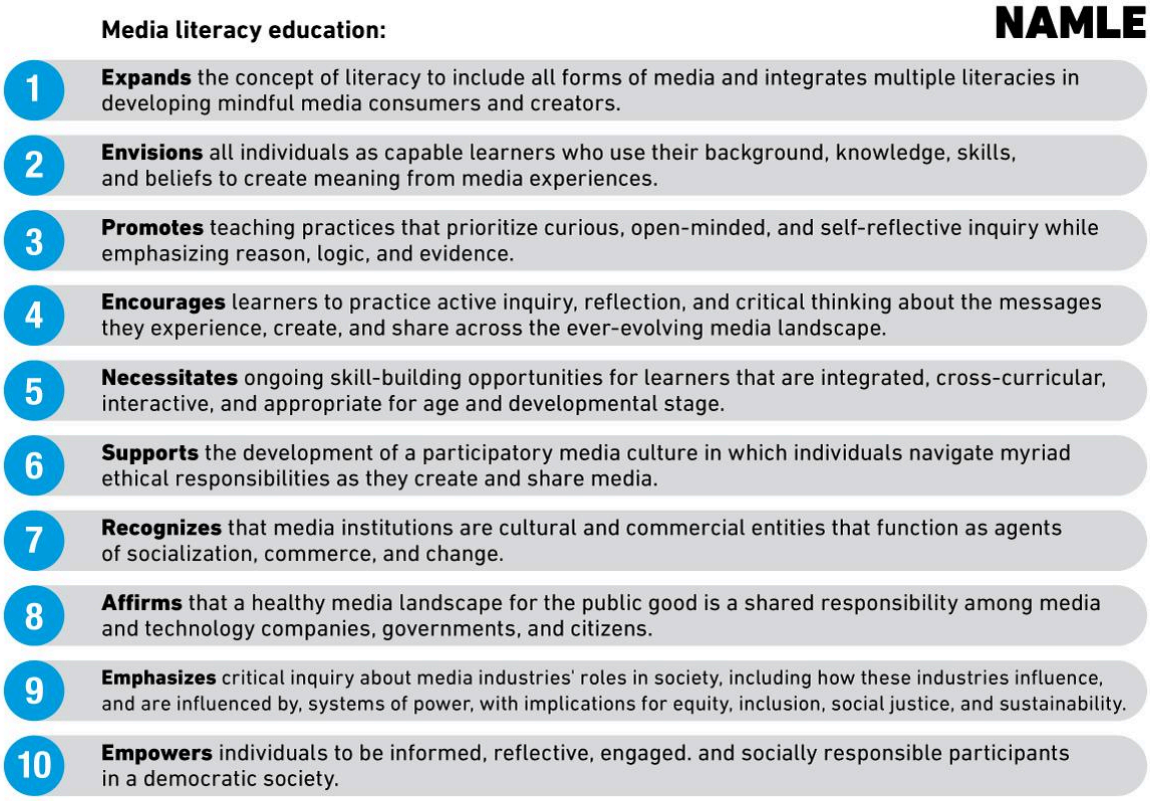
3.1.1. Examples of MIL initiatives

A plurality of actors has set forth initiatives to strengthen media literacy education. In recent years, they have placed a sharper focus on emerging technologies, such as AI. These initiatives aim to ensure that individuals are equipped with the knowledge and skills necessary to navigate this complex environment.

United States

In the United States, the notion of MIL and its core competencies have been advanced at the level of principles and non-binding recommendations. For example, the US National Association for Media Literacy Education (NAMLE), a non-profit organisation that serves as an umbrella for initiatives on media literacy education, drafted the 2007 Core Principles of Media Literacy Education, further revised in 2023. This non-binding document is a cornerstone of MIL in the USA, since the principles can be used as guidelines to implement policies and laws on enhancing media literacy education. The goal is to encourage the emergence of critical thinkers, effective communicators, and informed, responsible citizens, across all forms of communication channels, including computers, video games, radio, television, mobile media, print, and emerging technologies (NAMLE, 2023a).

NAMLE's Implications for Practice - an accompanying document to the Core Principles - provides concrete guidelines for implementation, exemplifying attitudes, values, teaching techniques, and classroom strategies that can best support each core principle (NAMLE 2023b). NAMLE's core principles are relevant to other countries seeking to introduce MIL in their educational strategies and have been translated into other languages.



Media literacy education:

NAMLE

1. **Expands** the concept of literacy to include all forms of media and integrates multiple literacies in developing mindful media consumers and creators.
2. **Envisions** all individuals as capable learners who use their background, knowledge, skills, and beliefs to create meaning from media experiences.
3. **Promotes** teaching practices that prioritize curious, open-minded, and self-reflective inquiry while emphasizing reason, logic, and evidence.
4. **Encourages** learners to practice active inquiry, reflection, and critical thinking about the messages they experience, create, and share across the ever-evolving media landscape.
5. **Necessitates** ongoing skill-building opportunities for learners that are integrated, cross-curricular, interactive, and appropriate for age and developmental stage.
6. **Supports** the development of a participatory media culture in which individuals navigate myriad ethical responsibilities as they create and share media.
7. **Recognizes** that media institutions are cultural and commercial entities that function as agents of socialization, commerce, and change.
8. **Affirms** that a healthy media landscape for the public good is a shared responsibility among media and technology companies, governments, and citizens.
9. **Emphasizes** critical inquiry about media industries' roles in society, including how these industries influence, and are influenced by, systems of power, with implications for equity, inclusion, social justice, and sustainability.
10. **Empowers** individuals to be informed, reflective, engaged. and socially responsible participants in a democratic society.

Figure 4. NAMLE Core Principles of Media Literacy Education (adapted from NAMLE, 2023)

Most US media literacy education policies are focused on school curricula and put emphasis on teaching students to identify false information. In California, the 2023 State Bill No.873 mandates the teaching of media literacy skills, including discerning false information, identifying misinformation, and promoting responsible internet content creation.

North Carolina stands out for being one of the first states to introduce specific guidelines on the ethical application of generative AI in public education. The guidelines comprise practical recommendations, with actions targeted at specific age groups and levels of maturity. From a young age, students critically analyse media, acknowledging that AI can manipulate images and videos. In middle school, students are encouraged to engage with AI, fostering AI literacy along with creativity, collaboration, and critical thinking. They are also made aware of the potential risks of AI misuse in social media and gaming. High school students receive specific training on large language models like ChatGPT, enhancing their understanding and responsible use of AI. They are informed about the potentially harmful applications of AI-generated content in social media and gaming.

At the federal level, bipartisan legislation was introduced in late 2023, named the Artificial Intelligence Literacy Act. This legislation aims to make AI literacy a key component of digital literacy education in schools, colleges, universities, and libraries. Although the document does not explicitly mention the implications of AI for mis- and disinformation, it aims to educate students about AI-related dangers, which encompass the dangers of AI-generated false information. The development of the act is based on partnerships with non-profit organisations, and its implementation must engage stakeholders and communities in all stages, in particular communities disproportionately impacted by the digital divide, including minority and rural communities.

European Union

In Europe, MIL principles were first enshrined in the European Charter for Media and Literacy. The charter was underpinned by three main tenets: critical, cultural, and creative (Frau-Meigs, 2024). The aim of the three 'Cs' is to ensure that EU citizens build their media literacy from concrete and culturally embedded experiences that help them develop their critical and creative thinking.

While the European Charter for Media and Literacy was a soft and non-binding document, the EU Audiovisual Media Services Directive (2018/1808), also known as AVMSD, requires member states to promote measures that develop media literacy skills (Art. 33.a). In the context of the directive, media literacy is defined as 'the skills, knowledge, and understanding that allow citizens to use media effectively and safely' (Recital 59). It should aim to equip citizens with the critical thinking skills required to analyse complex realities and recognise the difference between opinion and fact. The directive also recognises the need for cooperation with all relevant stakeholders, including media service providers and video-sharing platforms, to promote the development of media literacy.

In 2018, the EU Commission launched the Communication 'Tackling online disinformation: a European approach', which outlined key overarching principles to guide action on raising public awareness about disinformation. The communication engendered the 2018 EU Action Plan against Disinformation (European Commission, 2018a). One of the pillars of the Action Plan is dedicated to 'raising awareness and improving societal resilience'. It presented a structured approach to addressing issues requiring both reactive (debunking, fact-checking, and reducing the visibility of disinformation content) and proactive longer-term efforts (media literacy and measures to improve resilience).

In 2022, the European Union released Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training (European Commission, 2022a). The guidelines aim to promote the responsible and safe use of digital technologies and foster better public awareness and knowledge regarding disinformation, including AI-generated content. The guidelines provide pedagogical expertise and offer practical recommendations.

Throughout the years, the EU Commission has provided financial support to projects aimed at fighting disinformation. The European Digital Media Observatory, supported by the European Union, is a cross-border, multidisciplinary community of independent fact-checkers and academic researchers, who leverage their knowledge of local information environments to detect, analyse, and expose disinformation campaigns in Europe.

ASEAN

In response to growing concerns about the impact of social media on the information landscape, in 2018 ASEAN approved the Core Values on Digital Literacy, summarised in Table 1. This set of values was encompassed in a declaration and framework, signed by ASEAN ministers responsible for information, aiming to minimise the harmful effect of fake news.

Core values	Description
Responsibility	To think first and be responsible for what we post online

Empathy	To be respectful and thoughtful of how online interactions may affect others
Authenticity	To be sincere in our online interactions and prepared to stand by what we post
Discernment	To critically evaluate online information before acting on it
Integrity	To do the right thing, stand up for what is right and speak up against negative online behaviour

Table 1. ASEAN Core Values on Digital Literacy (ASEAN, 2018)

In 2022, ASEAN launched the Digital Literacy Programme (ADLP), a two-year initiative to fight mis- and disinformation in member countries. It aims to emphasise the role played by education in countering disinformation, including by proposing a toolkit to ‘train the trainers’, focused on building the capacity of secondary school teachers, university tutors and lecturers to empower their students in the fight against mis- and disinformation (ASEAN, 2022).

In 2024, AI was given a sharper focus in the region, and the ASEAN Guide on AI Governance and Ethics (2024) was launched. The guide emphasises the importance of safeguarding citizens against the use of AI for harmful purposes, such as generating or spreading disinformation. Governments are urged to enhance AI literacy, empowering citizens to fight misinformation from AI systems.

3.2. Content policy regulation and the role of intermediaries

In the digital context, intermediaries are mostly private agents, and they exist in the three layers of the internet: telecommunications infrastructure, technical standards, and applications. This means that the notion of intermediary may comprise various services, including internet service providers (ISPs), domain name system (DNS) service providers, cloud providers, search engines, e-commerce platforms, and social media platforms.

These intermediaries are important nodes in the network, enabling them to potentially act as gatekeepers, controlling the data and information flowing through their infrastructure.
In recent years, some actors have been seeking to put pressure on intermediaries to exert control over content. For example, ISPs have been requested to block access to certain websites or applications, and the Internet Corporation for Assigned Names and Numbers (ICANN) has been asked to block Russian country-code top-level domains (such as .ru) in order to counter disinformation in the context of the war in Ukraine (Fedorov, 2022).

Social media platforms also frequently receive take-down requests from government authorities. While the removal of some illicit material, such as child pornography, has become a consensual practice across intermediaries, the approach to address other types of content, especially involving speech, is much more difficult to harmonise.

The use of points of infrastructural control as proxies to gain control over internet flows carries some risks (DeNardis and Musiani, 2016). Internet intermediaries operating in lower layers of the internet - such as telecommunications and standards-setting organisations - carry out an important technical mission, necessary to ensure the integrity and interoperability of the internet. Co-opting these actors with the aim of achieving non-technical policy goals could lead to internet fragmentation.

The issue is more nuanced in the layer of applications, which involves websites and platforms. In particular, the development of platforms centred on user-generated content, such as social media, has led to a discussion on the limits of intermediary liability and the extent of obligations related to content removal.

Content policy often adopts a protective stance in relation to a myriad of different referent objects that need to be safeguarded (i.e. citizens, vulnerable segments of the population, democracy, public safety, or national security). At present, content policy has also become a go-to way of tackling mis- and disinformation, and within this scope, intermediaries play an important role.

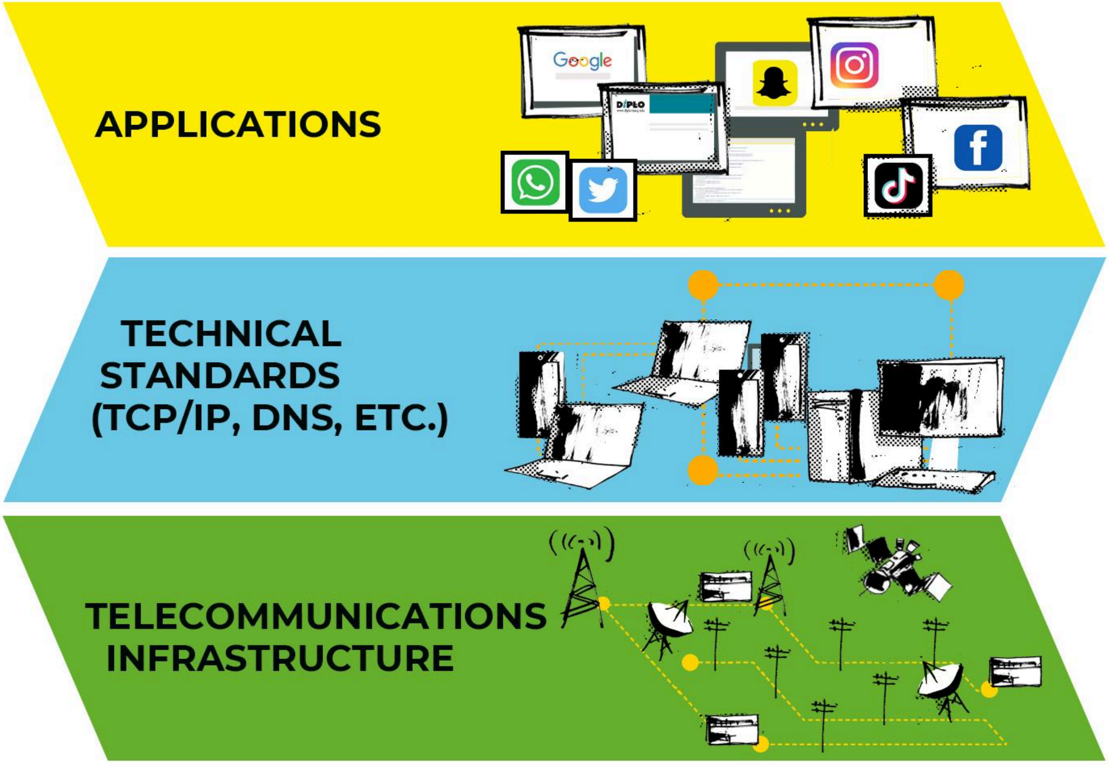

APPLICATIONS
Google
DIPLO
TECHNICAL STANDARDS (TCP/IP, DNS, ETC.)
TELECOMMUNICATIONS INFRASTRUCTURE

Figure 5. A multi-layered representation of the internet (DiploFoundation, undated).

3.2.1. Government initiatives to tackle mis- and disinformation

Online content policy is one of the areas that displays the largest number of regulatory and policy interventions introduced by governments (Evenett and Fritz, 2022). Between 2019 and 2020, at least 62 laws were proposed, amended, or implemented to tackle mis- and disinformation, but the upward trend could be noticed from 2015 (Yadav et. al 2021).

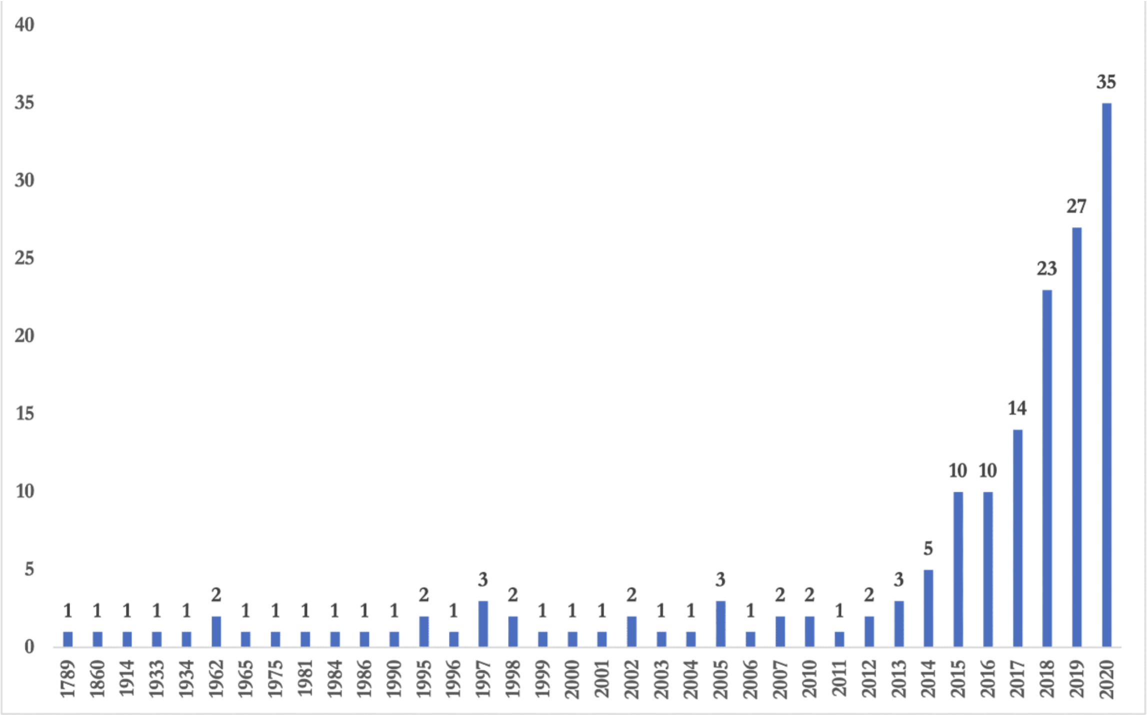


Figure 6. Laws to tackle mis- and disinformation over the last centuries (Yadav et. al 2021)

Some examples of policy and regulatory frameworks put in place by governments can be found below. Among them, the regulatory models of the United States and the European Union stand out. While the USA has set the first benchmark in terms of intermediary liability, the EU Digital Services Act, which is relevant in the context of fighting mis- and disinformation, has explicitly stated its extraterritorial application, becoming a global reference point. ASEAN provides an example of a soft and non-binding approach, based on enhancing convergence among member states.

United States

In the early days of the internet, John Perry Barlow penned A Declaration of the Independence of Cyberspace (Barlow, 1996). This anthology document, which reflected the libertarian internet culture at the time, was an ode to freedom of expression online, as well as a push-back against government intervention and regulation. There were high expectations that the internet would level the playing field, giving a voice to the powerless. Internet users largely relied on intermediaries to transmit, host, and share information with others online. As a consequence, these actors were seen as enablers of the right to freedom of expression and of the creation of an online public sphere, which would lead to the strengthening of democracy.

In this context, limiting the liability of intermediaries for content posted by third parties on their platforms was seen as a necessary measure to protect free speech and innovation. If intermediaries could be held liable for third-party content, they would have an incentive to take down potentially problematic content preemptively, in order to avoid litigation. This incentive could be especially strong in cases in which the definitions of illegal content are vague, or where it is not easy to determine whether the disputed content is unlawful (CDT, 2010). Without protection from liability, companies would also be less likely to invest in the development of new platforms, further entrenching incumbents in the social media segment.

This understanding helps explain the adoption of broad limitations to the liability of intermediaries in the USA, where most of the prominent companies that own social media platforms are based. Section 230 of the US Communications Decency Act significantly limits the civil liability of intermediaries. It states that providers or users of ‘interactive computer services’, which include internet service providers as well as platforms such as Facebook and X (former Twitter), cannot be treated as publishers of - and thus be held liable for - content produced by others. In parallel, it also allows those companies to voluntarily take actions in ‘good faith’ to take down objectionable material.

US federal courts have interpreted Section 230 as creating expansive immunity for claims based on third-party content. Consequently, internet companies not only frequently rely on Section 230 to avoid liability in federal and state litigation, but have also traditionally adopted a more hands-off approach in terms of content moderation in the USA. Since the USA is in the unique position of being home to many of the world’s leading tech companies, it has set the regulatory benchmark for how these companies would operate worldwide, by simply exercising ordinary domestic lawmaking (Chander and Sun, 2023).

Several other countries have enacted similar limitations to intermediary liability. The USA has pushed for the inclusion of provisions limiting intermediary liability within the trade agreements it has celebrated with third parties. The United States, Mexico, and Canada free trade agreement (USMCA FTA) is an example of this trend.

Although the model of Section 230 could be seen as one of the most protective of freedom of expression worldwide, it fails to strike a balance with the legitimate interests that could be harmed by the dissemination of online content (Moncau, 2020). This broad limitation to intermediary liability has created conflict in jurisdictions where certain specific types of speech are not allowed, such as hate speech and negationist speech related to certain historical events, such as the holocaust, in certain European countries (Rosenthal, 2020).
In recent years, there has been mounting criticism of section 230 in the USA, based on the understanding that courts have interpreted immunity too broadly.

Several bills aimed at amending the scope of Section 230 have been proposed in Congress (Brannon and Holmes, 2024). Scandals, such as the Cambridge Analytica affair and the alleged Russian interference in US elections, as well as news that social media has contributed to child exploitation and suicide (Yang, 2024), have rendered views on the impact of social media more negative in the USA.

There is also a growing understanding that the business model of social media platforms does entail some degree of control over the content displayed to users. Exercising control over content has become easier with algorithms. For example, AI is being used by social media platforms to identify and fact-check false content, and to remove illegal content, or content that violates the terms of use of the platform. With the alleged goal of preventing platforms from abusing this power, former President Donald Trump issued Executive Order 13925 on Preventing Online Censorship.

The EO directed federal agencies to cut their online ad spending with certain social media platforms and kickstart a federal rulemaking to reinterpret Section 230, aiming to limit platforms’ capacity to remove content in good faith. The EO was allegedly a retaliatory move against X for fact-checking claims made by former President Trump (Mackey, 2021), and an attempt to dismantle platforms’ efforts to curb the proliferation of false information (Allyn, 2020).

President Biden revoked Trump’s order, but some Democrats would like to repeal Section 230 altogether: they deem that this provision allows tech companies to avoid responsibility to do more in combating mis- and disinformation and hate speech online. In the US Congress and the judiciary, there are ongoing discussions about the limits of intermediary liability. In particular, the lack of clarity on how to deal with disinformation continues to raise legal issues. The ongoing *Murthy v. Missouri* case, for example, involves a claim that US federal government officials coerced social media companies to remove certain online content. The federal government contended that the content spread disinformation on healthcare issues, but the plaintiffs argued that it violated the right to freedom of expression enshrined in the First Amendment to the US Constitution.

European Union

EU efforts to combat disinformation date back to March 2015, and were initially focused on influence operations originating from abroad. In that year, the European Council invited EU member states and institutions, 'to develop an action plan on strategic communication to address Russia's ongoing disinformation campaigns'. This led to the creation of the strategic communications division (StratCom) and the first of its task forces East StratCom Task Force with a mandate to counter disinformation originating outside the EU. In 2017, two more StratCom task forces were created: one for the Southern Neighbourhood and another for the Western Balkans.

In 2018, an EU Action Plan against disinformation was launched, based on four pillars: 1. Improving the capabilities of EU institutions to detect, analyse, and expose disinformation; 2. Strengthening coordinated and joint responses to disinformation; 3. Mobilising the private sector to tackle disinformation; 4. Raising awareness and improving societal resilience (European Commission, 2018a). The 2018 Action Plan provided a foundation that has been expanded upon with the Code of Practice on Disinformation, published in 2018 and revised in 2022 (EU, 2022a), the European Democracy Action Plan (European Commission, 2020), and the Digital Services Act (EU, 2022b).

The Digital Services Act (DSA) applies to intermediaries ranging from e-commerce to social media, but puts more emphasis on Very Large Online Search Platforms (VLOSEs) and Very Large Online Platforms (VLOPs), with more than 45 million of monthly active users in the European Union. The DSA has introduced legal obligations to combat disinformation, as well as liability in the event of nonfulfilment of these obligations. Under the DSA, there are two main possibilities to combat disinformation:

- If the content is illegal - which refers to information that is incompatible with EU law or with the law of any member state, as per Article 3(h) - then all types of online intermediaries are obliged to act upon it.
- If the content is 'socially harmful', the DSA imposes specific obligations only on a selected category of online intermediaries (VLOPs and VLOSEs).

In the case of illegal content, platforms have several obligations, such as providing mechanisms enabling any person or entity to report content deemed illegal (Article 16), taking action against illegal content upon judicial or administrative order (Article 9(1)), and ensuring priority handling of notices submitted by entities referred to as trusted flaggers (Article 22).

VLOPs and VLOSEs are also obliged to prevent the dissemination of harmful content, which does not necessarily have to be illegal. These companies have the obligation to assess the systemic risks arising from the design, operation, and use of their services, as well as from the potential misuse of services (Article 34).

In the light of the DSA, disinformation may potentially entail systemic risks, particularly related to: a) actual or foreseeable negative impact on democratic processes, civic discourse, and electoral processes, as well as on public security (Recital 82); b) risk related to an actual or foreseeable negative effect on the protection of public health, minors and serious negative consequences to a person's physical and mental well-being, or on gender-based violence (Recital 83). In particular, these latter risks may stem from coordinated disinformation campaigns related to public health, or from online interface design that may stimulate behavioural addictions among service recipients (Recital 83). In consonance with the DSA, the European Commission (EC) sent requests for information on generative AI risks to six VLOPs and two VLOSEs (European Commission, 2024a). The EC requested these platforms to provide information on their mitigation measures for risks linked to generative AI, such as the spread of disinformation, deepfakes, and the automated manipulation of services.

According to the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, the transparency and due process requirements set forth by the DSA could help more to address the problem of disinformation, than a purely content-based approach, which seeks to restrict speech (UNGA, 2021).

In focus: the strengthened Code of Practice on Disinformation

The DSA is complemented by commitments encompassed within the framework of the strengthened Code of Practice on Disinformation (European Commission, 2022b). Adherence to the code is voluntary but may constitute proof of compliance with the obligations imposed by the DSA on VLOPs and VLOSEs. Among the signatures are large tech companies, such as Google, Meta, Microsoft, and TikTok.

The Code of Conduct includes the following commitments:

1. Demonetising advertisements containing disinformation
2. Labelling political advertising more clearly with details on the sponsor, advertising spending, and display period
3. Creating searchable databases of political advertisements
4. Empowering users to spot and flag non-factual information
5. Empowering fact-checkers, and strengthening their collaboration with platforms by integrating fact-checking services
6. Reducing malicious and manipulative behaviours used to spread disinformation (i.e. malicious deep fakes, bot-driven amplification, and fake accounts)
7. Ensuring that online services include safeguards against disinformation by design
8. Providing researchers and fact-checkers with better access to platforms' data
9. Making public the implementation efforts via a transparency centre
10. Adopting regular reporting and assessment mechanisms for the implementation of the code's commitments

Association of Southeast Asian Nations (ASEAN)

The views of ASEAN countries about the socioeconomic impact of the sharing of online information have changed in a short period of time. In 2014, the focus was on the positive impact of different types of media on access to information, education, and economic prosperity, as clearly stated in the Declaration on Social Responsible Media for a Peaceful and Prosperous ASEAN Community (ASEAN, 2014). At that stage, concerns related to harmful content had not yet been raised.

In 2018, ASEAN issued a framework aimed at minimising the harmful effects of fake news. This framework focused on the negative effects of false information and recommended using four main strategies to tackle the issue: education and awareness, detection and response, norms and guidelines, and community participation. While the framework gives governments the responsibility to oversee actions in these areas, it also acknowledges the importance of the involvement of non-state actors.

In 2023, Indonesia, which held the ASEAN chairmanship, published the ASEAN Guideline on Management of Government Information in Combating Fake News and Disinformation in the Media (Republic of Indonesia, 2023). The guideline provides a framework for governments' responses to the harmful effects of fake news and misinformation in the media and on social media platforms, seeking to establish standards and good practices, improve transparency and accountability in government communications, as well as improve coordination and collaboration between government agencies, especially at times of crisis. There is also a growing recognition that disinformation should be tackled in a multistakeholder way.

3.2.2. Frameworks of international organisations

One of the key roles of international organisations is to provide a platform for dialogue, where countries can share challenges and exchange good practices among a wide range of stakeholders. When discussions mature, international organisations develop guidelines and frameworks that can help synchronise and align national efforts to combat disinformation, while respecting fundamental rights and freedoms.

Another important role of international organisations is to facilitate research and data sharing on disinformation trends and impacts. By aggregating and disseminating research findings, they help stakeholders understand the evolving nature of disinformation and the effectiveness of different countermeasures.

United Nations: some highlights of the work conducted within the organisation

At the UN, disinformation is tackled by the UN Secretary General's office, Secretariat, and specialised agencies. UN initiatives include monitoring, analysing, and responding to the threat of mis- and disinformation to deliver United Nations mandates.

The issue of content policy is currently framed as 'information integrity' at the UN, and 'refers to the accuracy, consistency and reliability of information'. Information integrity is threatened by disinformation, misinformation, and hate speech. (United Nations, 2023). Disinformation must be deterred, as it represents a threat to scientifically established facts, and to the realisation of SDGs (United Nations, 2024). Examples of the latter include gender-based hate speech and disinformation that seeks to undermine action against climate change.

At the same time, the UN General Assembly and the Human Rights Council have recognised that responses to the spread of disinformation should comply with international human rights law and promote, protect, and respect the right of individuals to freedom of expression, including the freedom to seek, receive, and impart information (UNGA, 2021). The UNGA has also acknowledged the importance of investing in prevention and resilience to disinformation through media literacy initiatives (UNGA, 2022).

In 2022, the UN Secretary General published the report 'Countering disinformation for the promotion and protection of human rights and fundamental freedoms' (United Nations, 2022). The report describes challenges and threats related to disinformation, sets out the relevant international legal framework and discusses the practices taken by states and businesses to counter disinformation. In 2023, the Secretary General published Policy Brief 8 on information integrity on digital platforms (United Nations 2023). The ideas contained therein were further developed within the 2024 publication ‘Global Principles for Information Integrity’.

In focus: The UN Global Principles for Information Integrity

The UN Global Principles provide a framework for multistakeholder action against disinformation. The document advances five principles, underpinned by respect for human rights, and provides a set of recommendations.

- **Societal trust and resilience**. In this context, trust refers to the confidence that people have in the sources and reliability of the information that they access, including official sources and information, and in the mechanisms that allow information to flow throughout the ecosystem. Resilience refers to the ability of societies to handle disruptions or manipulative actions within the information ecosystem.
- **Healthy incentives.** This includes addressing the critical implications for information integrity resulting from business models that depend on targeted advertising and other forms of content monetisation. The framework calls for a fundamental shift in incentive structures, and for the adoption of human rights-driven business models that do not rely on algorithm targeted advertising based on behavioural tracking and personal data.
- **Public empowerment.** People should have control over their online experience, should be able to make informed decisions as to the media they choose to consume, and should express themselves freely. Public empowerment requires consistent access to diverse and reliable sources of information. It also requires tech companies to enhance user control and choice, including interoperability with a range of services from different providers.
- **Independent, free, and pluralistic media.** A free press is a cornerstone of information integrity and democratic societies. There are ongoing threats to press freedom, such as online and offline harassment of media workers, as well as the migration of advertising revenue to the digital space, dominated by large tech companies. Robust responses should include support to public interest news organisations, journalists and media workers, and media development assistance.
- **Transparency and research.** Increased transparency by tech companies and other information providers can enable a better understanding of how information is spread. Ensuring privacy-preserving data access for a diverse range of researchers helps to fill research gaps and inequalities. Academics, journalists and civil society must be protected and supported in carrying out their vital work free from fear or harassment.

UN Human Rights Council and UN Special Rapporteur on freedom of opinion and expression

The UN Human Rights Council and UN Special Rapporteurs have helped to clarify the negative impact of disinformation on human rights, while, at the same time, shedding light on the way that policies to combat disinformation may curtail human rights and fundamental freedoms (see Topic 1.1). In particular, the impact that measures to combat disinformation may have on the right to hold opinions and on the right to freedom of expression have been mainstreamed.

In the 2021 report on disinformation, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, pointed out that the responses by states and companies have been problematic, and detrimental to human rights (UNGA, 2021). While state responses have often been 'heavy handed', companies also play a major role in spreading disinformation, 'but their efforts to address the problem have been woefully inadequate'. It is possible to distil some remarks and suggestions from the report, such as:

- Enhancing clarity when it comes to the definition of disinformation adopted by domestic laws. Adopting different definitions and approaches to tackle mis- and disinformation is particularly necessary.
- Discussions on disinformation should encompass the problem of state-sponsored disinformation, which can emanate from state institutions directly or from proxies targeting audiences within the state's own territory or abroad for political and strategic aims. When states systematically and simultaneously suppress other sources while promoting their own false narratives, they are denying individuals the right to seek and receive information under the ICCPR.
- Placing greater emphasis on the way that States and tech companies may undermine the right to freedom of opinion. Punishment and manipulation of people's opinions are practices to be avoided by public and private actors.
- Freedom of expression can only be restricted by a narrowly construed law, under the cases foreseen by Article 19(c) of the ICCPR - the legitimate aim of respecting the rights and reputations of others and protecting national security, public order, public health, or morals. The scope, meaning, and effect of the domestic law need to be sufficiently clear, precise, and public. The directness of the causal relationship between the speech and the harm, and the severity and immediacy of the harm, are key considerations in assessing whether the restriction of speech is necessary.
- The principle of necessity requires the restriction to be appropriate and proportionate to achieve the legitimate aim, using the least restrictive means to protect it. This means that criminal sanctions constitute serious interference with the freedom of expression and are disproportionate responses in all but the most egregious cases.
- Political speech should undergo a high threshold of legality, legitimacy, necessity, and proportionality.
- Countries should not delegate to companies the role of 'speech police', which may lead to the suppression of legitimate online expressions with limited or no due process, and without prior court order. Regulatory proposals should focus on platforms' transparency and due process obligations.

The UN Special Rapporteur urged states to recalibrate their responses to disinformation, enhancing the role of free, independent and diverse media, investing in media and digital literacy, empowering individuals, and rebuilding public trust (UNGA, 2021).

United Nations Educational, Scientific and Cultural Organization (UNESCO)

UNESCO's approach to countering disinformation is to promote information as a public good. It does this through three main pillars: advocating for greater transparency on social media platforms, supporting independent journalism and educating audiences with critical thinking through media and information literacy (UNESCO, 2022a).

In 2019, UNESCO sponsored the Addis Ababa Declaration Journalism and Elections in Times of Disinformation (UNESCO, 2019). The declaration emphasises the critical role of free and independent journalism in supporting democratic elections, particularly in times of disinformation. The declaration calls for the creation and implementation of legal frameworks that ensure freedom of expression and privacy, the protection of journalists, and the avoidance of broadly worded regulations that might unduly restrict expression. It calls for multistakeholder dialogues to address challenges related to disinformation.

In 2021, UNESCO initiated a global dialogue to enhance the transparency of internet companies, with the release of 26 high-level principles, some of which involve recommendations to counter online disinformation. In 2023, UNESCO launched an action plan to regulate social media platforms that seeks to strike a balance between regulation and protecting freedom of expression and human rights. The plan offers guidance on platform governance, following a series of worldwide consultations backed by a global multistakeholder opinion survey.

UNESCO's action plan is based on **seven principles**, recommending that:

1. The impact on human rights becomes the compass for all decision-making, at every stage and by every stakeholder;
2. Independent, public regulators are set up everywhere in the world with clearly defined roles and sufficient resources to carry out their mission;
3. These independent regulators work in close coordination as part of a wider network, to prevent digital companies from taking advantage of disparities between national regulations;
4. Content moderation is feasible and effective at scale, in all regions and all languages;
5. Accountability and transparency are established in these platforms' algorithms, which are too often geared towards maximising engagement at the cost of reliable information;
6. Platforms take more initiative to educate and train users to think critically;
7. Regulators and platforms take stronger measures during particular sensitive moments like elections and crises.

UNESCO also assessed the problem of disinformation during elections, particularly in the publication Elections in Digital Times: A Guide for Electoral Practitioners (UNESCO, 2022b). This guide serves as a resource for electoral practitioners, including electoral management bodies. It addresses the impact of social media, digital communication, and emerging technologies on election campaigning, information sharing, and opinion shaping. It highlights the challenges of disinformation and emphasises the need for structural solutions, and digital literacy.

In 2024, UNESCO issued a policy brief titled User empowerment through media and information literacy responses to the evolution of generative artificial intelligence (GAI) (UNESCO, 2024). The document argues that the risks posed by mis- and disinformation in the context of generative artificial intelligence should be addressed through MIL, promoting individual empowerment. The policy brief calls for a shared global vision in formulating and implementing public policies to empower people through AI/GAI. This vision could be developed and sustained by some institutional innovations, such as the establishment of a global Media and Information Literacy (MIL) observatory, tasked with funding and producing evidence-based research about the impact of AI literacy on well-being, education, and society. It also advocates the creation of a UN Oversight Body on Information and AI, with the involvement of all stakeholders, for regular monitoring and reporting on MIL.

World Health Organization (WHO)

The interplay between mis- and disinformation with other public policy areas, such as health, climate change, and security, is becoming increasingly clear. During the COVID-19 crisis, the term 'infodemic' (Rothkopf, 2003) rose to prominence, as the World Health Organization (WHO) used it as a metaphor to describe a scenario of excessive information (both real and false), in which mis- and disinformation spread like a virus. Examples could be noticed during the COVID-19 crisis, and included attempts to downplay the pandemic and suggest that it was a hoax. Mis- and disinformation represented obstacles to governmental disease control strategies, and lowered individuals' intent to receive vaccination, as exemplified by a study conducted with individuals in the United States and the United Kingdom (Loomba et al., 2021).

Some of the WHO's key initiatives and collaborations on infodemics include:

- A pilot programme called Early AI-supported Response with Social Listening (EARS) that uses artificial intelligence to provide real-time insights into online discussions surrounding COVID-19. This programme helps health authorities understand public opinion and respond effectively to concerns.
- An initiative called 'Verified' aimed at providing fact-based verified content on COVID-19.
- Partnership with the Wikimedia Foundation to expand access to reliable information about COVID-19. This collaboration makes trusted public health information available on Wikimedia Commons, a digital library of free images and resources.
- Work with social media platforms like Facebook, Google, and Twitter to promote reliable information and counter misinformation. These platforms have implemented measures to limit the spread of false information and provide accurate updates on the pandemic.
- Development of a policy framework for managing the infodemics. WHO organised a few events addressing misinformation.

3.2.3. Initiatives of non-governmental actors

Tech companies and civil society play a relevant role in content governance. Their initiatives vary from cooperation with governments in multistakeholder frameworks to companies' self-regulation and civil society-led initiatives, such as the International Fact-Checking Network (IFCN), established by the Poynter Institute.

Tech companies

For tech companies and platforms, the question of content is highly sensitive, and they are under contradictory pressure on this issue. On the one hand, the use of their platforms for the dissemination of mis- and disinformation often triggers strong public reactions and media coverage. As a consequence, tech companies are interested in containing this risk for their operations and businesses. On the other hand, platforms seek to maximise ad revenue, hence algorithms are designed to prioritise eye-catching content, which may include sensationalist posts, clickbait, and disinformation.

Companies' initiatives to mitigate the problem of mis- and disinformation have been centred around:

1. Fact-checking initiatives: Many companies have implemented fact-checking tools and partnered with organisations that verify the accuracy of information shared on their platforms. This helps in identifying and flagging false or misleading content. Facebook, for example, is working with nearly 100 third-party fact-checking organisations, which review and rate the accuracy of content on Facebook. If they identify false information, Facebook reduces the distribution of those posts - but does not remove them entirely, as a way to mitigate potential issues related to freedom of expression.

There are also initiatives implemented in specific countries. For example, Google has developed partnerships with fact-checking organisations like the Australian Associated Press (AAP) to increase the speed and quantity of fact-checks and distribute them to users in Australia and New Zealand. Google has also launched features like 'About this image' and 'About this result' to provide users with more context about the sources of information they encounter online.

2. Transparency measures: Companies are focusing on transparency by requiring advertisers to confirm their identities, disclosing information about political advertising, and enabling users to see who has placed and paid for adverts. This transparency helps build trust and accountability.

3. Content moderation: Platforms have content moderation systems in place to remove harmful or misleading content that violates their policies. This helps reduce the spread of disinformation on their platforms. There are also initiatives to enhance the transparency of decisions related to content policy. For example, Facebook's Oversight Board, which reviews content moderation decisions, seeks to enhance transparency and accountability. By adhering to these self-imposed guidelines, companies inspire to build trust with users and stakeholders.

4. Educational campaigns: Companies are investing in educational campaigns to promote information literacy among users. Google's 'Be Internet Awesome' programme offers interactive lessons and activities that teach children a basic knowledge of digital citizenship and safety. Similarly, YouTube has partnered with the Poynter Institute's MediaWise project to improve digital literacy among young users. This collaboration focuses on teaching users how to identify reliable sources and fact-check information. Facebook's 'We Think Digital' initiative targets various demographics, promoting digital literacy skills. Through partnerships with third-party fact-checkers, Facebook also educates users about the significance of fact-checking and how it can help maintain the integrity of the information they consume.

5. Partnerships and alliances: Collaborating with fact-checkers, news organisations, and civil society groups is another way companies address disinformation. By working together, they can develop effective strategies to combat false information.

6. AI and technology: Companies are leveraging artificial intelligence and technology to detect and filter out disinformation. AI tools can quickly analyse vast amounts of data, identifying patterns that human moderators might miss. It is also important to exercise caution and avoid overreliance on AI without adequate human oversight. AI systems can sometimes misidentify legitimate content due to their limited understanding of context and nuance. Good practices involve combining AI with human judgement to ensure more accurate content moderation. Facebook's use of its Oversight Board to review content decisions made by AI systems helps balance automated efficiency with human discernment, facilitating the appropriate handling of difficult cases.

One important technological approach is the use of **watermarks in AI-generated content**. Watermarking involves embedding a unique identifier in digital content, making it easier to trace the origin and authenticity of the material. This technology can help identify and verify AI-generated images, videos, and texts, thereby mitigating the spread of disinformation. The EU Artificial Intelligence Act (European Union, 2024) and US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House, 2023) include provisions on transparency measures, ensuring that AI-generated content is clearly marked and traceable

7. User empowerment: Platforms are focusing on empowering users to report and flag misleading content. By providing users with tools to report disinformation, companies involve the community in the moderation process. Social media platforms like Facebook and Twitter have integrated features that allow users to report content they believe is false or misleading. These reports are then reviewed by a team of moderators or fact-checkers who assess the validity of the claims. Facebook prioritises the evaluation of content that has been flagged multiple times, ensuring that potentially harmful disinformation is addressed promptly. Twitter has introduced a community-based approach to misinformation called Birdwatch. Birdwatch allows users to identify information in Tweets they believe is misleading and write notes that provide informative context.

Companies are adopting a multifaceted approach that combines technology, partnerships, transparency, and user education to address the issue of disinformation effectively.

Trusted flaggers

Trusted flaggers emerged in response to the growing challenge of managing the vast amounts of user-generated content on digital platforms. As early as the 2000s, online platforms struggled with the influx of content that needed moderation. Initially, moderation was handled in-house by platform employees, but as content volume surged, these platforms began to enlist the help of external parties working as trusted flaggers. Trusted flaggers are individuals or organisations with expertise in identifying harmful or inappropriate content. They were introduced to help platforms manage content more effectively and efficiently by flagging posts that violate community guidelines or legal standards. These flaggers, due to their demonstrated expertise and reliability, were given priority in the review process.

As digital platforms grew and the complexity of content moderation increased, so did the scope and sophistication of trusted flagger responsibilities. Platforms began to formalise these programmes, offering training and tools to trusted flaggers to improve the accuracy and efficiency of their reports. Nevertheless, as noted by Maréchal et al. (2020), platforms' algorithms may target users who are most vulnerable to disinformation in order to cater to their interests, while hiding disinformation-related content from other users who would otherwise be in a position to flag it and provide corrective counter speech. This means that platforms' automated content optimisation may become an obstacle to the work of trusted flaggers.

The EU's Digital Services Act (DSA), has strengthened the role of trusted flaggers. The DSA has institutionalised and standardised the role of trusted flaggers across the EU, providing a framework for their operation. Under the DSA, trusted flaggers are recognised as one of the key stakeholders in the content moderation ecosystem. The act outlines specific criteria for becoming a trusted flagger, ensuring that only individuals or entities with proven expertise and reliability are granted this status. These criteria include:

1. **Expertise:** Trusted flaggers must demonstrate significant experience and expertise in identifying and reporting harmful content.
2. **Independence:** They must operate independently of the platforms they assist, ensuring unbiased reporting.
3. **Accountability:** Trusted flaggers are required to adhere to high standards of accuracy and reliability, with mechanisms in place to review their performance regularly.

The DSA mandates that online platforms prioritise and respond promptly to reports from trusted flaggers. This prioritisation is designed to expedite the removal of harmful content, enhancing user safety and platform accountability. The DSA also encourages greater transparency by requiring platforms to publish regular reports on their content moderation activities, including the role and impact of trusted flaggers.

Civil society and human rights organisations

Civil society organisations (CSOs) play an important role in combating mis- and disinformation. In 2021, the National Endowment for Democracy (NED) launched a working paper that mapped civil society organisations that work specifically to combat disinformation (Bradshaw and Neudert, 2021). It identified initiatives from 175 CSOs across Africa, Asia, Europe, Latin America, and North America.

The research classified the work of civil society into six different non-mutually exclusive categories: (1) credibility initiatives; (2) verification initiatives; (3) education and media literacy; (4) research and tool provisions; (5) developing norms, standards and policy recommendation; and (6) initiatives to support journalism.

Civil society organisations have also brought attention to the interplay between laws on disinformation and freedom of expression. For example, the Center for International Media Assistance tracked the impact of laws on fake news in a 2023 report. Between 2011 and 2022, 78 countries passed laws designed to limit the spread of false or misleading information on social media.

The publication noted the downsides of laws that focus on content, criminalising the creation and distribution of fake news. Since the definition of fake news is vague, governments have discretionary power to define what content is prohibited, generating a chilling effect on freedom of expression. As a consequence, laws on mis- and disinformation are often constraining press freedom. Improving platform transparency and accountability, along with providing media and digital literacy, could be a better course of action.

Every year, Freedom House publishes 'Freedom on the Net', a globally recognised survey that assesses the state of internet freedom. The 'Freedom on the Net 2023' report discusses the issues of mis- and disinformation, particularly highlighting the impact of AI-generated content and acknowledging that AI has significantly increased the scale and sophistication of disinformation campaigns. Moreover, disinformation campaigns, particularly those involving non-consensual deepfakes, disproportionately target women and vulnerable groups. These attacks aim to damage the reputation and discourage public participation by these groups, exacerbating existing societal inequalities.

The report calls attention to the responsibility borne by both governments and private sector actors in spreading disinformation. State-backed campaigns often hire private firms or influencers to covertly manipulate online information. This outsourcing provides governments with plausible deniability and complicates the attribution of disinformation efforts.

4. Initiatives to counter disinformation in the context of elections

In the specific context of political processes, disinformation can distort public perception and influence electoral outcomes, undermining the right to free and fair elections—a cornerstone of democratic governance. Disinformation may also weaken democratic participation. The Parliamentary Assembly of the Council of Europe (PACE) expressed special concern with widespread information pollution, disinformation campaigns, and foreign electoral interference in Resolution 2326 (PACE, 2020).

In 2024, the global stage is set for a plethora of elections, with at least 83 slated worldwide, as reported by the New York Times (Hsu et al., 2024). The rapid development of technology, especially AI, has put social media platforms in the spotlight, as they will play an important role in political campaigns. In its Global Risks Report 2024, the World Economic Forum identified misinformation and disinformation as the most significant global risk over the next two years, particularly affecting countries scheduled to hold elections (WEF, 2024). According to the report, this dissemination of false information could compromise the integrity of electoral outcomes and government legitimacy. The importance of digital communication for elections creates the potential for abuse. An emblematic example is the alleged interference by the Russian government in the 2016 United States elections with the goal of spreading disinformation and influencing the electoral process.

AI may influence how information is gathered and opinions are formed. An emerging concern is related to AI's capability to create deepfakes — manipulated images, videos, or audio files that appear strikingly realistic. Their potential to falsely depict public figures carries implications for political life.

During the election campaign in Slovakia, for example, glimpses of the disruptive potential of video-based deepfakes could be noticed. Michal Simecka, a progressive member of the European Parliament found himself embroiled in a scandal fabricated by AI - a video circulated on social networks purportedly showed him in conversation with a journalist, discussing how his party had purchased votes from the Roma minority (Meaker, 2023). Simecka was quick to denounce the video. Slovak authorities had already issued warnings against the proliferation of disinformation engineered using AI. While it is difficult to assess whether the deepfakes manipulated Slovak voters' choices—and to what extent—it became clear that artificial intelligence is increasingly being used to target elections and could disrupt future ones (Levine and Savoia, 2023).

Audio-based deepfakes are also causing concern. A report from the Center for Countering Digital Hate (CCDH, 2024) investigated the vulnerabilities and risks associated with AI voice-cloning tools in the context of elections. The report tested six popular AI voice-cloning tools. In 80% of 240 tests, these tools successfully generated convincing audio statements of high-profile politicians. These false statements included claims about corruption, election fraud, bomb threats, and health scares.

The study found that none of the tools had adequate safeguards against misuse of their technology. Only one of the tools succeeded in blocking the voice cloning of US and UK politicians, but failed with EU politicians. The report also documented the real-world usage of AI voice cloning for disinformation in elections in the USA, UK, Slovakia, and Nigeria. The study concluded that existing election laws should be updated to address the new risks posed by AI-generated content, and that voluntary commitments by AI companies are insufficient.

In parallel, policies initially aimed at fighting disinformation may be easily misused and abused by public authorities allowing governments greater control and discretion over public discourse, imposing arbitrary or politically motivated limits to freedom of expression (APC, 2021).

Given the fundamental importance of freedom of expression to democracy and the enjoyment of all other human rights and freedoms, international human rights law affords particularly strong protection to expressions on matters of public interest, including criticism of governments and political leaders and speech by politicians and other public figures. According to the Special Rapporteur, any restriction on disinformation in the context of political speech requires a high threshold of legality, legitimacy, necessity, and proportionality. Electoral laws may justifiably forbid the propagation of falsehoods relating to electoral integrity, but such a restriction must be 'narrowly construed, time-limited and tailored so as to avoid limiting political debate' (UNGA, 2021).

4.1. Government-led initiatives to combat mis- and disinformation in the context of elections

In response to the potential threats exacerbated by AI, the US Federal Election Commission (FEC) established detailed regulations in 2022, requiring clear disclaimers indicating who is responsible for campaign content. These rules were developed to foster transparency in online political advertising, and address both text/graphic and audio/video online political communications

In 2023, the state of Michigan passed a law against AI deepfakes during elections (Cappelletti and Swenson, 2023). Minnesota, California, Washington, and Texas, already have laws restricting AI use in political communications, with the goal to prevent the spread of mis- and disinformation (Ahmed, 2023). Hawaii has introduced a bill against AI-generated deepfakes and disinformation in political campaigns ahead of the 2024 elections (DeJournette, 2024). The bill aims to hinder the spread of political mis- and disinformation by prohibiting the distribution of electioneering communications before an election that a person knows or should have known are deceptive and fraudulent deepfakes of a candidate or party. States like Wisconsin, Florida, and New York also have pending legislation, although their approval in time for the upcoming elections remains uncertain.

In the UK, one of the key features of the Elections Act 2022 is the introduction of digital imprints for political campaigning. This measure was designed to increase the transparency of online political campaigning by requiring campaign materials disseminated digitally to include an imprint stating who is behind the campaign content. This aims to provide voters with comparable levels of transparency between offline and digital political campaigning.

The EU has been active in tackling online misinformation through a range of initiatives. These include implementing regulations, such as the Digital Services Act (DSA) promoting media literacy programmes, and establishing the European Digital Media Observatory to monitor and counter misinformation online.

The EU Parliament has adopted rules focussing on online political advertising requiring clear labelling and prohibiting the sponsoring of ads from outside the EU (European Parliament, 2024). The European Commission has issued guidelines under the DSA, targeting VLOSEs, in order to protect the integrity of elections from online threats (European Commission, 2024b). The new guidelines emphasise tailored risk mitigation and collaboration with authorities and civil society. The proposed measures recommend companies to establish internal teams, conduct elections-specific risk assessments, adopt specific mitigation measures linked to generative AI, and collaborate with EU and national entities to combat disinformation and cybersecurity threats. The platforms are urged to adopt incident response mechanisms during elections, followed by post-election evaluations to gauge effectiveness.

EU political parties have also signed a code of conduct brokered by the European Commission intending to maintain the integrity of elections (Griera, 2024). The signatories pledge to ensure transparency by labelling AI-generated content and abstain from producing or disseminating mis- and disinformation.

4.2. Private-led initiatives on political advertising to combat mis- and disinformation in the context of elections

In recent years, major social media platforms have faced intense scrutiny and criticism over their handling of political advertisements. In response, they have introduced new policies aimed at increasing transparency and curbing the spread of mis- and disinformation during electoral cycles.

In 2018, Meta (the parent company of Facebook, Instagram, and Threads) introduced the Ad Library, a publicly accessible database that stores details of all active and inactive ads related to social issues, elections, or politics as of March 2019. The database aims to provide greater advertising transparency by making a comprehensive and searchable collection of all advertisements running across platforms including Facebook and Instagram, readily available to the public. In February 2024, Instagram also announced that Instagram and Threads would stop recommending political content from accounts that users do not already follow (Instagram, 2024). The company, however, did not clarify what ‘political’ means in the context of this decision.

In 2019, Twitter banned political advertising, arguing that political influence should be earned through genuine interest rather than purchased amplification. However, in January 2023, Twitter announced that it would begin relaxing its ban on political ads, allowing advocacy groups and elected officials to resume promotions focused on specific causes (Conger, 2023).

During the 2020 US presidential elections, social media platforms took various measures to combat disinformation and election interference. For instance, Twitter used warning labels for misleading posts, while YouTube removed videos with false claims about election fraud. Facebook and WhatsApp implemented strategies to curb the spread of fake news and limit the forwarding of messages.

In 2024, a coalition of 20 major tech companies, including OpenAI, Microsoft, Adobe, TikTok, and X, launched a joint initiative to combat deceptive AI content potentially threatening global elections (Dang and Paul, 2024). They committed to collaborating in several areas, such as developing content identification tools, public awareness campaigns, and measures against inappropriate content on their platforms. Potential measures under discussion include watermarking or embedding metadata to certify the origin of AI-generated content. As part of this initiative, Google introduced limitations on its AI chatbot, Gemini, restricting it from answering questions related to elections in the countries where they were held (Robins-Early, 2024).

Private companies have also been tailoring potential solutions to the EU political landscape. TikTok has launched an Election Centre within its app, tailored to EU member states’ languages, to combat misinformation in the context of elections (Morgan, 2024). The aim is to detect and remove content conveying disinformation and covert influence campaigns by collaborating with local electoral commissions, civil society groups, and fact-checking organisations. Additionally, the company aims to focus on misleading AI-generated content by requiring content creators to label this type of content. Meta has also announced a team to tackle the spread of disinformation and the misuse of generative AI in the lead-up to the European Parliament elections in 2024, addressing potential threats in real time (Yun Chee, 2024).

Google's Jigsaw Unit, dedicated to addressing societal threats, has announced the launch of a series of animated ads across platforms such as TikTok and YouTube (Coulter, 2024). The videos convey 'prebunking' techniques, helping viewers recognise manipulative content before it gets widely disseminated.

5. Case studies

5.1. Finland: a media literacy champion

Finland has taken a proactive stance on combating mis- and disinformation and fostering a resilient information environment. The country has been particularly recognised for policies and educational strategies aimed at equipping citizens with the critical thinking skills necessary to navigate a complex media landscape (Mackintosh, 2018). The country's approach to countering mis- and disinformation is multifaceted, encompassing education, media literacy, and strategic communication.

Finland does not have specific laws that criminalise the spread of disinformation, but rather addresses the issue through various measures, with emphasis on media literacy, transparency, and free press protections. Existing laws related to defamation, hate speech, and national security may indirectly address certain aspects of disinformation. In 2019, Finland was ranked first out of 35 countries in a study of resilience to the post-truth phenomenon (Lessenski, 2018).

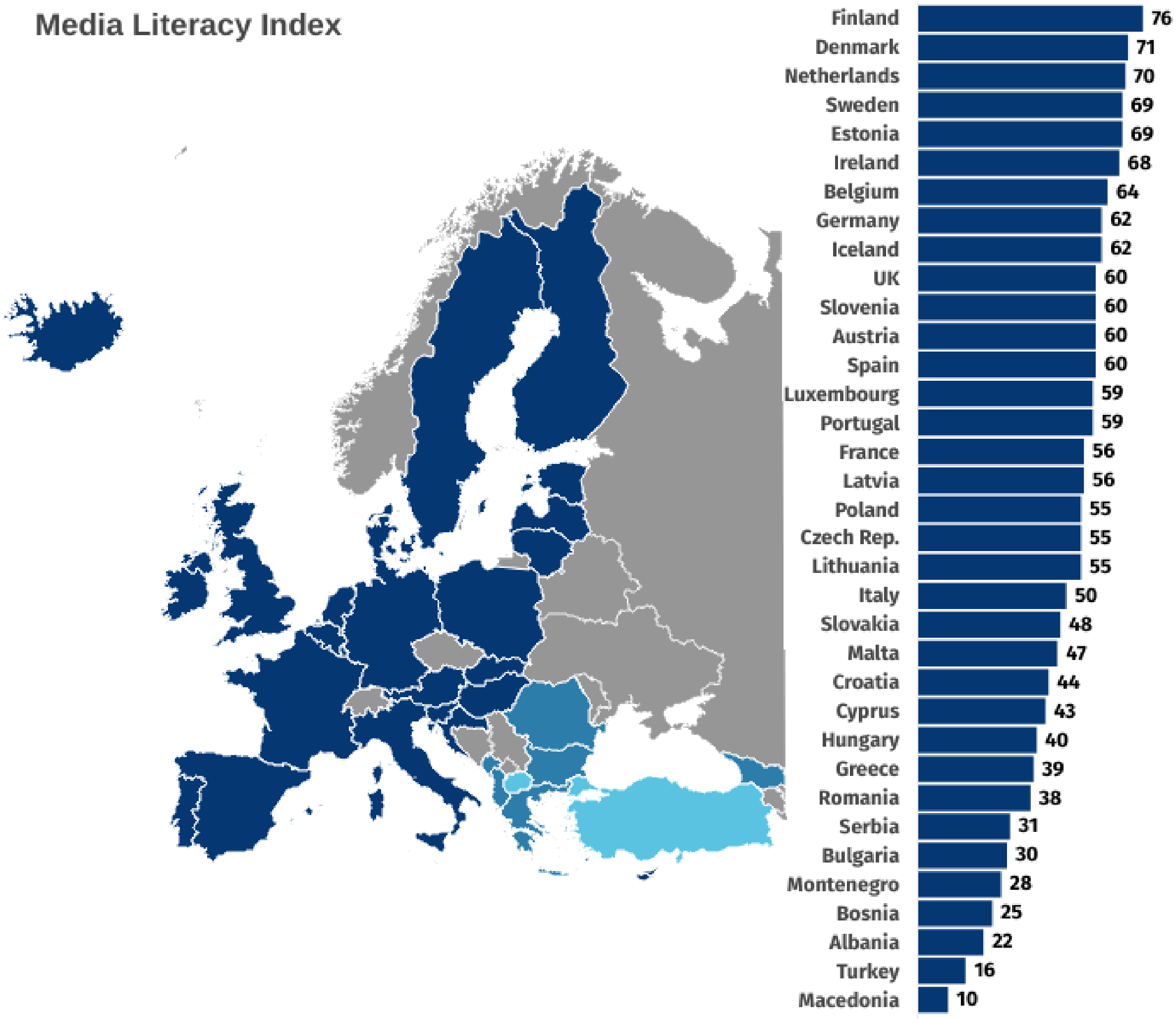


Figure 7. Media literacy across Europe (adapted from Pettersson, 2018)

As a consequence, other countries have started looking to Finland as a model for combating mis- and disinformation (Mackintosh, 2018), which enhances the importance of studying the Finnish approach to tackling the issue. Moreover, Finland's collaborative efforts between government agencies, educational institutions, and the media play a vital role in fostering a well-informed public. The Finnish government actively engages with media outlets to ensure transparency and accuracy in reporting while promoting public awareness campaigns that highlight the dangers of mis- and disinformation. These initiatives are supported by robust policies encouraging accountability and integrity in information dissemination.

Initiatives to combat mis- and disinformation

Following Finland's independence from Russia in 1917, the country faced Kremlin-backed propaganda campaigns (Kivinen, 2022). This resulted in the implementation of policies that tackle such propaganda and the spread of mis- and disinformation that could threaten national security.

In August 2022, the Finnish Security and Intelligence Service (Supo) proposed to criminalise the dissemination of information by actors acting on behalf of foreign states with the aim of influencing Finnish decision-making maliciously (YLE News, 2022). In order to mitigate the impact on freedom of expression, the focus was on individuals who knowingly spread disinformation campaigns. The proposal faced resistance from actors wishing to put in place a precise legal definition of the problem as a first step (YLE News, 2022). Public debate and media literacy were still seen as primary defences against disinformation.

On 4 April 2024, Finnish Foreign Minister Elina Valtonen and US Secretary of State Antony Blinken signed an MoU at the NATO Ministers meeting to counter foreign state information manipulation (Finnish Ministry of Foreign Affairs, 2024). Under the MoU, countries are expected to develop comprehensive strategies beyond mere 'monitoring and reporting', such as investing in technical and human capacity and to designate specific governance structures and institutions to oversee national efforts.

The MoU recognises that civil society, independent media, and academia play crucial roles in supporting government initiatives to fight disinformation. Additionally, multilateral organisations are deemed vital for international cooperation, helping to address information and capability gaps among partner nations.

Media and literacy

According to the Open Society Institute's Media Literary Index, Finland ranked first among 41 European countries (Open Society Institute, Sofia, 2022). The National Media Education Policy, jointly published by the Ministry of Education and Culture and the National Audiovisual Institute in 2019, delineates Finland's approach to media literacy considering the growing importance of Information and Communications Technologies (ICTs).

The 2019 curriculum encompassed mis- and disinformation in the context of traditional media, incorporating the study of propaganda, advertising, and misleading statistics (EDMO, undated). Finland introduced its first formal media and information literacy curriculum in 2004, focusing on addressing media violence (OECD, 2023).

The Finnish media education framework involves collaboration with diverse stakeholders. For instance, non-governmental partners like civic organisations, schools, libraries, NGOs, and universities play a crucial role in contributing to the comprehensive nature of Finland's media education initiatives.

The National Audiovisual Institute (KAVI), under the Ministry of Education and Culture, is mandated to advance media education. Through its Department for Media Education and Audiovisual Media (MEKU), KAVI collaborates with stakeholders to promote media literacy and a safer media environment for children. KAVI coordinates Finland's national media education policy, operating primarily at the national level while supporting regional initiatives. It aids regional and local organisations in strategic media education planning, reinforcing the implementation of media literacy across the country. Additionally, KAVI focuses on developing and promoting media education practices, models, and pedagogies. It enhances educators' media literacy awareness and competencies, offering support through resources like the online Media Literacy School.

One of the resources published by the Media Literacy School is *Explore the phenomenon: Misinformation.* This portal offers guidance for teachers on how to educate students between the ages of 13-17 to identify mis- and disinformation. This includes practical exercises to understand when their dissemination is intentional or not. The exercises aim to foster critical thinking and the ability to evaluate the reliability of online information sources. The exploration of deepfake videos is another resource targeted at individuals between 13-17 years old. The resources focus on educating individuals to understand what a deepfake is, how to recognise it, and its wider impact.

Although Finland has been praised for media literacy education focused on the education system and the youth, there are fewer initiatives targeted at the adult population. This gap raises concerns about how the country would handle a major coordinated disinformation effort (Hyvärinen, 2024). Between 2015-2016, the limits of Finland’s media literacy efforts became apparent, when a significant domestic disinformation campaign was produced against the affluence of a record number of asylum seekers (Hyvärinen, 2024). Websites pretending to be legitimate news sources spread false stories and racist messages, which many Finns shared on social media. This incident demonstrated a susceptibility to disinformation, but it was homegrown rather than foreign-led. The campaign exploited people's tendency to believe information aligning with their personal views, indicating that even in a nation with strong media literacy initiatives, there are vulnerabilities when adults are not equally educated on these issues.

Reflections

The Finnish approach to combating mis- and disinformation offers a model of resilience in the information environment. The country is recognised for its comprehensive policies and educational strategies aimed at equipping citizens with critical thinking skills essential for navigating a complex media landscape. Finland’s strategy focuses on media literacy education, transparency, and maintaining free press protections without specific laws criminalising disinformation.

In 2019, Finland was ranked first out of 35 countries in terms of resilience to the post-truth phenomenon, demonstrating effective measures against misinformation and the sway of emotion over facts. Its proactive stance includes collaboration between government agencies, educational institutions, and media outlets, promoting transparency and accuracy in reporting. Public awareness campaigns emphasise the dangers of misinformation, supported by robust policies ensuring accountability in information dissemination.

While Finland's media literacy education has been globally praised, it predominantly targets the youth, leaving gaps in adult education. Nevertheless, Finland’s approach, involving multilateral engagements and comprehensive national strategies, serves as a valuable reference for other nations in the fight against disinformation, balancing the protection of freedom of expression with the need for information integrity.

When it comes to countering disinformation, Finland's strengths are a high level of institutional and media trust, as well as teaching media literacy which is essential in the development of critical thinking.

5.2. Sweden: strengthening trust and social resilience against external threats

Like other countries in Europe, Sweden has a long trajectory of policies to combat disinformation, dating back to the Cold War. Since the 1950s, bodies aimed at countering disinformation, propaganda, and information operations have been put in place. The year 2022 represents a watershed, marked by the creation of the Psychological Defence Agency (PDA), with a clear mandate to identify and counter disinformation in partnership with public institutions and other stakeholders in society. The PDA has become a reference for other countries, especially in the Nordic region.

The Swedish approach to combating disinformation provides a relevant case study. First, while some countries spread out their attention, and adopt general frameworks that could help them combat any type of disinformation perceived as harmful, Sweden places particular emphasis on combating disinformation originating from abroad, focusing on protecting national security and democratic order. In parallel, there is a deep-rooted commitment to preserving freedom of opinion and expression among Swedish citizens, preserving the vibrancy of democratic debate.

Another distinguishing characteristic of the Swedish approach to countering disinformation is the existence of a central body. The PDA leads efforts to coordinate the operations of other agencies, promoting a whole-of-government approach. The agency adopts a two-pronged emphasis on promoting situational awareness of threats and national capabilities on the one hand, while strengthening resilience at a societal level, on the other. In the context of the latter, media literacy strategies are given prominence.

Sweden does not have laws against disinformation, but there are laws against defamation, inciting ethnic hatred, agitation and sedition, which could be applied against certain types of disinformation. Sweden's national security strategy seldom mentions information influence operations as a threat on their own, but recognises the importance of psychological defence, crisis preparedness, and civil defence in building the country's resilience to various threats, stating that 'influence campaigns and disinformation via digital platforms undermine people's trust in government agencies and impact our security' (Government Offices of Sweden, Prime Minister's Office, 2024).

The Psychological Defence Agency

In Sweden, certain administrative tasks are delegated to public agencies. These agencies operate independently within their areas of expertise, implementing laws and policies and delivering public services. Government instructions set overarching goals for the agencies, but they formulate concrete objectives and strategies on their own.

The PDA was created in January 2022, amid growing concern about Russian influence (Lee Myers, 2023). The agency is tasked with identifying, analysing, and providing support in ‘countering malign information influence and other misleading information that is directed at Sweden or Swedish interests by antagonistic foreign powers’ (PDA, undated). Malign information influence can be defined as an ‘attempt to harmfully influence, disrupt or steer public discourse in Sweden’, carried out by foreign powers or other external threat actors (PDA, undated).

In order for information to be considered malign, it does not need to be incorrect, but it must be deceptive (Kozłowski, 2024). If an adversarial foreign actor exercises control over the information environment, this actor may become able to convey an unbalanced and lopsided account of events. This one-sided version, regardless or being correct or not, impinges on people’s capacity to form their own opinions and may characterise an attempt to disrupt or steer public discourse in Sweden.

Democratic principles forbid the PDA to monitor domestic actors. This means that in order for the PDA to act, the influence attempt must come from outside of Sweden, and present a clear intention to undermine the Swedish government or citizens (Kozłowski, 2024). Sweden places high importance on the capacity of citizens to exercise freedom of opinion and expression. In practice, however, external information influence campaigns can be sometimes difficult to separate from legitimate domestic opinion, especially in the context of astroturfing, which can be understood as the practice of hiding the sponsors of a message or organisation to make it appear as if it originates from, and is supported by, grassroots participants.

Actions developed by the PDA have the main goal of strengthening defence, especially through boosting psychological resilience. Psychological defence refers to society’s common ability to detect and resist mis- and disinformation directed at Sweden. Strengthening the spirit of resistance puts emphasis on the attitude of the population and the capacity of individuals to resist disinformation in times of peace or war. The willingness to defend is underpinned by society’s trust in the state. Building trust in peacetime is seen as a fundamental prerequisite for identifying and countering disinformation, as well as for creating the conditions to effectively respond in times of crisis (Timm, 2022). This particularly resonates with the stance taken by the UN Global Principles for Information Integrity, which underscores the importance of societal trust and resilience (United Nations, 2024).

Although the PDA is mainly focused on defence, it also has an offensive mandate, to the extent that the agency is expected to prepare Sweden for war and psychological warfare (Kozłowski, 2024). If Sweden is at war, or at risk of being at war, the PDA will provide support to the government with advice and capabilities to counter any aggressor’s intent to attack the country.

In carrying out its actions, the agency places emphasis on collaborating with the media, which plays an important role in society’s ability to manage disinformation crises. Upon request, the PDA supports media companies in identifying, analysing, and responding to undue influence on information. Freedom of the press, free media, and freedom of expression are seen as preconditions for psychological defence.

Collaboration with academia is also important in Sweden. The Psychological Defence Research Institute (PDRI) of Lund University, partially funded by the PDA, has been particularly active in promoting discussions and sponsoring publications. The latest includes the report Building Resilience and Psychological Defence (Palmertz et al., 2024). The report provides a practical analytical guidebook and a toolkit, which could be useful to other countries considering the implementation of a domestic strategy to combat disinformation.

The framework is built on three main pillars: assess, address, and evaluate, as presented in Figure 8.

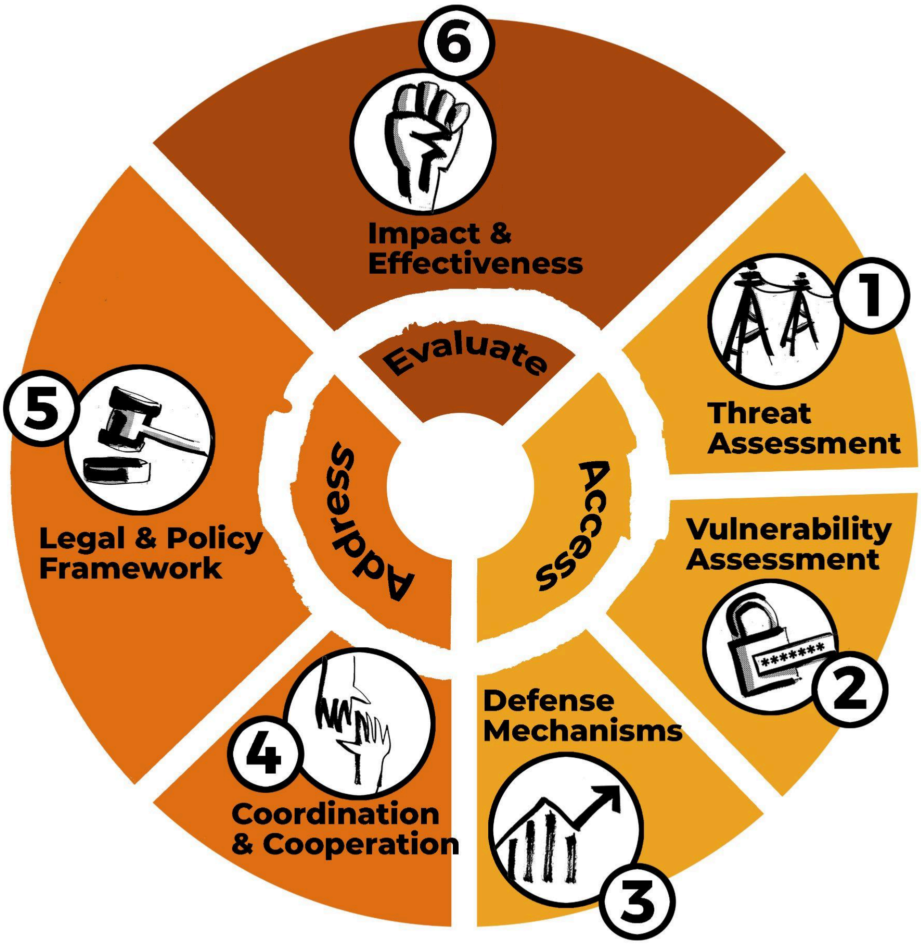


Figure 8. The analytical framework for countering hybrid threats and foreign influence and interference (adapted from Palmertz et al., 2024).

- **Assess** refers to the mapping of *external* threats - denoting antagonistic actors that may seek to exercise malign influence and the *internal* vulnerabilities that these actors may seek to target, as well as the available defensive mechanisms.
- **Address** denotes the state's existing *capabilities* for addressing the threats and vulnerabilities identified, such as national coordination, international cooperation and existing legal and regulatory frameworks.
- **Evaluate** refers to an *integrated analysis*, with a view to establishing a holistic understanding of the impact of threats and effectiveness of capabilities identified above.

The PDA also supports research and analysis on emerging trends, as exemplified by reports that analyse the use of AI large language models (LLMs) in foreign information influence operations (Bjurling et al., 2024), and a publication analysing malign foreign interference and information influence on video game platforms (Falkheimer et al., 2023).

Media and information literacy

The Swedish approach to media literacy places emphasis on strengthening resilience at a societal level. Societal resilience is individual capacity at scale, and can be understood as the symbiosis of a group of individual capacities coming together (Fee, 2021). In this context, society is understood in a broader sense. Sweden adopts a whole-of-society approach to combating disinformation, which can be associated with the concept of total defence (Sörensen and Pamment, 2023). The entire country should be prepared to resist an attack, defend the country, and contribute to recovery efforts in the event of a crisis or conflict. This approach necessarily entails the collaboration between government and non-governmental actors, including the population in general, civil society organisations, journalists, and the private sector.

The PDA should strengthen the population's resilience, conduct training, promote cooperation, and ensure coordinated action in countering threats towards Sweden (Government Offices of Sweden, Ministry of Defence, 2021). In recent years, the PDA has increased its role in media literacy activities. To develop psychological defence and support actors in this endeavour, the agency has created an education and training structure. Activities include not only sponsoring publications, but also the provision of training courses. Education and training are seen as vital to countering disinformation because of the ripple effect these actions may have on the entire society. Courses and training provided by the PDA include knowledge of democratic principles and awareness of information influence operations conducted by foreign powers targeting Sweden, as well as tools for identifying and countering these threats (Sörensen and Pamment, 2023).

Between 2021 and 2023, the PDA delivered twenty-five basic courses, two implementation courses, and one advanced course, covering concepts such as ‘public awareness building’, ‘analysis and identification capabilities’, ‘strategic communication capabilities’, and some activities relating to ‘system-wide capabilities’ (Sörensen and Pamment, 2023).

Reflections

In Sweden, problems related to information disorder are framed as malicious information influence operations, which include, but are not limited to disinformation. The focus is not on whether information is correct or incorrect, but on the undue control of the information environment by foreign actors, which may lead to one-sided and biased information. This scenario is damaging to freedom of opinion and may lead to attempts to influence, disrupt or steer public discourse in Sweden.

The focus on external sources of IIOs has defined the realm of the PDA’s mandate. The agency remains independent from the Swedish government and does not monitor domestic actors. Its actions must be grounded on the constitutional right to freedom of expression. This shows that framing the problem of disinformation as a national security issue does not necessarily need to entail government control over the domestic information environment and censorship.

Compared to other countries, the Swedish approach puts more emphasis on a proactive approach, based on strengthening trust and collaboration between government bodies and citizens, as well as on a constant assessment of threats and capabilities. Instead of focusing on a reactive approach, which seeks to shape exogenous elements in the information environment, there is an understanding that Swedish society is able to cultivate societal resilience. This is achieved not only through actions aimed at media literacy at the individual and societal levels, strengthening trust between government and citizens, but also through monitoring the information environment, in order to identify early attempts to exploit vulnerabilities in Swedish society.

5.3. Lithuania: the importance of civic engagement

Lithuania's policies against mis- and disinformation are primarily focused on protecting national security, largely due to ongoing challenges from Russian propaganda (Mays, 2023). Over the years, Lithuania has adopted a multifaceted strategy that includes media literacy programmes, cybersecurity initiatives, and cooperation with international partners. This approach not only aims to enhance national security but also aims to build resilience within society against the influence of false information.

Examining Lithuania's case is crucial because civic engagement plays an essential role in deterring mis- and disinformation. Active engagement of civil society, independent media, and volunteer groups enhances public resilience and supports comprehensive anti-disinformation strategies.

Initiatives against mis- and disinformation

Lithuania addresses the issue of disinformation under its Law on the Provision of Information to the Public. Article 2 (13) defines disinformation as intentionally disseminated false information, whilst Article 19(2) prohibits its dissemination by stating that ‘it shall be prohibited to disseminate disinformation and information which is slanderous and offensive to a person or which degrades his honour and dignity’.

Additionally, according to Lithuania’s National Security Strategy, media literacy programmes are imperative for strengthening the country’s resilience. Article 37.6 correlates society's resistance to misinformation and other informational threats with the enhancement of the education system, critical thinking media and information literacy programmes, carried out in a collaborative manner between government and non-governmental actors.

The Lublin Triangle and the fight against propaganda

Russia’s invasion of Crimea prompted neighbouring countries to cooperate in the fight against disinformation. On 28 July 2020, the foreign ministers of Poland, Lithuania, and Ukraine established the Lublin Triangle, a tripartite cooperation rooted in the historical ties among the three countries (Ukrainian Ministry of Foreign Affairs, 2020). This initiative aims to enhance political, economic, and security collaboration.

The Lublin Triangle's primary goals include countering disinformation and enhancing societal resilience. The three countries aim to create a strong defence against hybrid threats by combining resources and expertise, ensuring regional stability and security. In 2021, they signed a Roadmap for expanding trilateral cooperation, focusing on countering hybrid threats and disinformation (Republic of Poland, 2021). A Joint Action Plan for 2022-2023 was developed to combat disinformation and strengthen resilience.

On 6 December 2022, during the EU-Ukraine Forum on Countering Disinformation in Brussels, three NGOs from the Lublin Triangle countries— Lithuania's Civic Resilience Initiative, Poland's Kościuszko Institute, and Ukraine's Detector Media—presented a joint report (Civic Resilience Initiative et al, 2021). This report highlighted the challenges of Russian disinformation and propaganda in the Lublin Triangle region and exemplified the close cooperation between the three partners.

The role of civic engagement

Lithuania has examples of successful civic initiatives against disinformation. In 2021, DELFI 'Melo Detektorius' (or 'lie detector', in English) an independent fact-checking unit affiliated with DELFI, Lithuania's largest internet news portal, was considered the best fact-checking success story in Europe.

DELFI's main goal was to identify and expose a Russian propaganda network and troll farm spreading disinformation about the Baltic States and the West to a large number of followers (Mays, 2023). Its fact-checking methodology involves selecting facts from meetings, public comments, and press conferences, and categorising findings as 'lie', 'partial lie', 'partial truth', 'truth', or 'manipulation' based on traceable evidence. Editorial pieces and speculative future events are excluded from their assessments. This project prompted Facebook to invite DELFI representatives to showcase the Lie Detector at the Virtual Global Summit.
Projects like Debunk.org, Lithuanian Elves, and the Civic Resilience Initiative have mobilised

Lithuania's media and citizens to debunk falsehoods, with an emphasis on fostering international civic collaboration to combat disinformation. These initiatives encourage the creation and dissemination of accessible, user-friendly tools designed for diverse demographics, including the youth, the elderly, and ethnic communities.

Lithuanian Elves

The Lithuanian Elves are a grassroots movement comprising thousands of volunteers dedicated to combating Russian disinformation and propaganda online, particularly on social media. Formed in 2014 in response to the Russian invasion of Crimea, in Ukraine, and the subsequent influx of pro-Kremlin propaganda targeting the Baltic states, the Elves have become a significant force in information warfare.

Named as a counterforce to the pro-Kremlin 'trolls', the Elves actively monitor online content, debunk false stories, and report accounts that spread disinformation. Their tactics include coordinating efforts to flag fake news comments under online articles, running 'blame and shame' campaigns against pro-Russian trolls, and exposing Russian disinformation to bolster societal resilience. They also manually check content that AI systems might miss to ensure accuracy. The movement has expanded from its initial 50 members to over 5,000 volunteers across Lithuania and the Baltic region.

Depending on the situation, the Elves may act proactively or reactively, operating individually and as a well-organised community. They check suspicious content, publicly debunk false stories, and call out websites and accounts spreading disinformation. As the adversaries have become more sophisticated, targeting societal weak points in the Baltics, the Elves have forged strong alliances with media outlets. Debunk.org is crucial in strengthening the volunteer community by organising training events and workshops, equipping the Elves with the necessary skills and tools. Their tireless efforts ensure a robust process, effectively complementing automated AI systems in the ongoing battle against disinformation.

Civic Resilience Initiative (CRI)

The Civic Resilience Initiative (CRI) is a Lithuanian non-profit, non-governmental organisation founded in Vilnius in 2018. Established by a group of experts from across Europe, CRI focuses on enhancing the resilience of Lithuanian and regional societies through engaging educational activities.

The organisation targets key areas such as security, media literacy, disinformation, and cyber issues, aiming to empower civil society to actively participate in educational efforts. CRI brings together experts to provide insights and fill educational gaps where government institutions fall short.

The Initiative collaborates with partners such as the NATO Public Diplomacy Division, the Konrad Adenauer Foundation, and Lithuanian ministries of foreign affairs and defence. Through these efforts, CRI promotes democratic processes and long-term societal resilience against various threats.

CRI has been issuing reports as guidelines in the fight against mis- and disinformation. These reports serve as crucial tools, outlining strategies and best practices to combat the spread of false information in various domains. By disseminating these guidelines, the CRI aims to empower organisations and individuals worldwide to effectively address and mitigate the harmful impacts of misinformation through informed actions and policies.

One of them is a report on the Challenges of the Contemporary Disinformation (CRI, 2020), followed by training targeted at journalists. The training aimed to enhance critical thinking and digital resilience against false information. The report highlights the role played by digital platforms in the rapid dissemination of harmful content, and the importance of challenging the platform's underlying business model, which relies on ad revenue.

Reflections

Several lessons emerge from Lithuania's efforts against mis- and disinformation. First, the emphasis on national security underscores the severity of threats posed by disinformation, particularly from state-sponsored sources like Russian propaganda. This necessitates a proactive approach combining legal frameworks, media literacy initiatives, and cybersecurity measures to safeguard democratic processes and societal resilience.

Second, Lithuania highlights the pivotal role of civic engagement in countering misinformation. Initiatives such as the Lithuanian Elves and the Civic Resilience Initiative showcase the effectiveness of grassroots movements and NGO collaborations in combating false narratives and enhancing digital literacy. These efforts empower citizens to actively participate in defending against disinformation, fostering a more informed and resilient society.

Moreover, Lithuania's involvement in regional alliances like the Lublin Triangle exemplifies the benefits of international cooperation in addressing hybrid threats and disinformation campaigns. By pooling resources and expertise across borders, countries can strengthen collective defence and uphold information integrity on a broader scale.

5.4 Singapore: seeking balance

Singapore ranks first in regional digital capabilities in Southeast Asia, followed by Malaysia (Cheng and Chow, 2023). The country has developed its Cyber Security Agency (CSA) along with key laws on cybersecurity, personal data protection, and computer misuse. In spite of that, Singapore lags behind in terms of ratifying international human rights conventions (OHCHR, undated).

Freedom House (2024) considers Singapore a partly free country due to its parliamentary system dominated by the ruling People's Action Party (PAP). When assessing internet freedom, more specifically, Freedom House (2023) also classified Singapore as partly free, largely due to government-enabled laws regulating online behaviours that threaten internet freedom in the country. This is reflected in online content policy, including on the way the country fights disinformation.

As early as in 2019, the Singapore government proposed a law specifically combating mis- and disinformation. After a series of public hearings about the proposal, a parliamentary committee concluded that the phenomenon of deliberate online falsehood creates a serious problem for Singapore, with an impact on the country's national sovereignty and security, social cohesion and democratic institutions. In May 2019, the Protection from Online Falsehoods and Manipulation Act (POFMA) was adopted by the parliament.

Initiatives to combat mis- and disinformation

Singapore adopted the term Hostile Information Campaigns (HICs) to make reference to deliberate attempts by foreign actors, often secretive and coordinated, to create and spread information to manipulate public opinion and harm the country's interests. HICs' tactics may range from destabilising the target country through inciting or inflaming social tensions, manipulating public opinions on sensitive issues, or undermining the public trust in the country's institutions. (Singapore Ministry of Home Affairs/MHA, 2022).

The government is adopting a two-pronged approach to protect Singapore from foreign interference and HICs: 1. enhance the legal framework to counter HICs and 2. educate citizens about the threat of disinformation and influence campaigns.

	Initiatives	**Explanation**
Legal/ regulatory framework	Protection from Online Falsehoods and Manipulation Act (POFMA), 2019	To safeguard against the spread of falsehoods via electronic means (Singapore MHA, 2022)
	Foreign Interference (Countermeasures Act (FICA), 2021	To strengthen Singapore's ability to prevent, detect and disrupt foreign interference in its domestic politics conducted through HICs and the use of local proxies. (Singapore MHA, 2022)
	Online Safety (Miscellaneous	To regulate egregious content within

	Amendments) Act, 2022	online communication services accessible to Singaporean users (Singapore MCI, 2023).
	Online Criminal Harms Act (OCHA), 2023	To enable the government to deal more effectively with online activities that are criminal in nature (Singapore MHA, 2023).
	Infocomm Media Development Authority's Online Safety Code, 2023	To mitigate the risks of harmful social media content to Singapore users, especially children, by requiring social media services to enhance online safety in Singapore. This safety code is legally binding (Singapore IMDA, 2023).
Education	Source, Understand, Research, Evaluate (S.U.R.E) Campaign, 2013	Rolled out by the National Library Board (NLB) Singapore to campaign the use of four concepts (S.U.R.E) when assessing news (Singapore MHA, 2022).
	Factually	Real-time updates on government websites and social media accounts to clarify common misperceptions that can harm Singapore's social fabric.(Singapore MHA, 2022).

Table 2. Singapore initiatives in combating mis- and disinformation (Sources: MHA, IMDA, MCI, and Freedom on the Net)

Protection from Online Falsehoods and Manipulation Act (POFMA)

POFMA employs the term 'false statements', not mis- and disinformation. Based on the law, a false statement is false or misleading, whether wholly or in part and whether on its own or in the context in which it appears. Commenting on the bill, the Ministry of Law emphasised that the law would target only falsehoods, not opinions, criticism, satire, or parody.

The criticism presented to this definition, based on the difficulty of telling apart facts and false statements in some cases, was dismissed by the government (Academia Singapore, 2019). In POFMA, the minister is authorised to distinguish facts from false statements. It is possible to appeal the decision, first to the minister who has passed the decision, and when the minister denies it, to the high court.

The government claimed that the act was aimed at serving the public interest, defined by POFMA as: a) ensuring the security of Singapore; b) protecting public health, finances, public safety or public tranquillity; c) maintaining friendly relations of Singapore with other countries; d) preventing any influence on Singapore electoral processes; e) preventing incitement of hatred; f) preventing a diminution of public confidence in the performance of public institutions.

POFMA also provides extraterritorial jurisdiction as it covers statements communicated in Singapore or to individuals located in Singapore through the internet or MMS/SMS. Several types of internet intermediaries - from access providers to platforms hosting third-party content - may be held liable according to the law. The potential sanctions under POFMA include heavy fines and imprisonment, in the case of individuals (Singapore Legal Advice, 2022).

In 2021, after two years of implementation, the International Commission of Jurists, - a non-government organisation established in 1952 and formed by lawyers and judges defending human rights - produced a briefing on the implementation of POFMA (ICJ, 2021). The commission highlighted the noncompliance of POFMA with international law and standards protecting the right to freedom of opinion, expression, and information due to the overly broad provisions, including the wide discretion given to the executive to define what is false or not, and the heavy penalties that include imprisonment and fines. The commission noted that most of the restrictions were directed at politicians outside the government and ruling party, government critics, and independent media outliers.

In 2023, Amnesty International (2023) expressed concerns about the implementation of POFMA, which may lead to persecution of human rights defenders and government critics. Although the problem of online falsehoods is real, POFMA may give the government authority to control content circulating online (RSF, 2019).

Media literacy in Singapore

Education is an important approach to combat mis- and disinformation in Singapore (MHA, 2022). Since 2011, the National Library Board (NLB) has been promoting information literacy by launching the National Information Literacy Programme (NILP), particularly targeted at young students in primary and secondary schools. The Applied Learning Programme (ALP) in media literacy provides another example of a project aimed at fostering critical thinking, to enable students to identify false information and media bias (Singapore Ministry of Education, undated). The project adopts a learning-by-doing approach, based on real-world examples and hands-on projects.

In 2023, the NLB introduced the S.U.R.E (Source, Understand, Research, Evaluate) campaign, aimed at heightening public awareness about the importance of critical evaluation skills when searching for information. S.U.R.E. is targeted at a broader audience, which includes students, teachers, parents, and the general public.

In terms of AI literacy, the Ministry of Education (MOE) of Singapore seeks to develop foundational knowledge of AI and promotes its safe, ethical, and responsible use in schools and Institutes of Higher Learning (IHLs) (Singapore MOE, 2024a). Educational curricula also incorporate cyber wellness programmes aimed at teaching students to critically assess information, detect fake news (including AI-generated), and grasp the fundamentals of data security, privacy, and responsible online behaviour (Singapore MOE, 2024b).

The Ministry of Communications and Information (MCI) launched the Digital Readiness Blueprint in June 2018, outlining 10 recommendations on how to help Singaporeans embrace technology (Singapore MCI, 2018). The learning outcomes within the framework provide a common guide that can be distilled into specific learning objectives, including topics of deepfakes, fake news, misinformation, and disinformation. The values that should be upheld during online activities, according to the blueprint, are similar to those proposed by Singapore in 'Core Values on Digital Literacy for ASEAN', adopted in 2018. As a result of these efforts, young people in Singapore have confidently assessed that their own digital literacy skills are the highest in ASEAN (UNICEF, 2021b).

Reflections

Singapore places national security as a core concern when developing regulations to combat mis- and disinformation. According to POFMA, the government has discretion over defining false statements, based on a broad understanding of what is considered to be a ‘threatened public interest’. Sanctions within POFMA apply not only to social media platforms but also to individuals.According to analysts, POFMA can be used to put pressure on human rights defenders and government critics, especially during political events such as elections.

In parallel, Singapore has strongly advocated for digital literacy, domestically and also in ASEAN. The government has fostered media literacy initiatives in the education system, including AI and disinformation. The continued efforts have made the level of digital literacy among young people one of the highest in ASEAN.

6. Combating disinformation: key takeaways

6.1 The need for a multidimensional approach

The complexity and scale of disinformation requires a comprehensive approach. In digitalised societies, regulation may be carried out not only by top-down laws enacted by governments, but also by the simultaneous influence of four main mechanisms: laws, norms, market, and technological architecture (Lessig, 2006). Laws are explicit and binding mandates produced by lawmakers that can be enforced by governments; norms are social conventions that one often feels compelled to follow, which include non-binding agreements, frameworks, and principles; market forces regulate by acting upon supply, demand and the pricing system and; the 'code' or architecture of the internet – the software instructions and protocols that underpin its functioning – constrains what can and cannot be done online (Lessig, 2006).

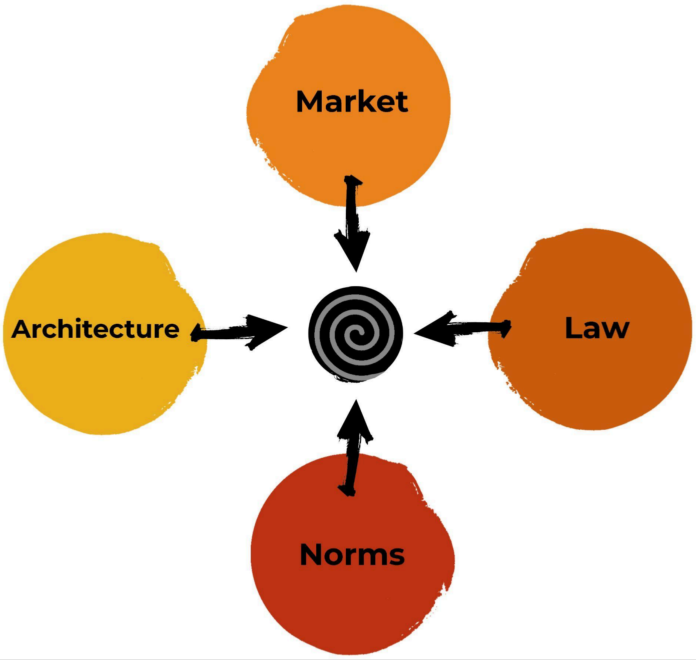


Figure 9. Lessig's socioeconomic theory of regulation (adapted from Lessig, 2006)

Laws

Governments have adopted different regulatory approaches to tackle problems related to information disorder. Some, such as Singapore, have enacted specific laws on disinformation, while others, such as Lithuania and Sweden, tackle the problem through existing laws that may apply to disinformation. More recently, there has been a significant trend of perceiving problems related to disinformation as threats to national security and to the democratic order, placing the issue under the umbrella of foreign information influence operations. This trend can be perceived across the four case studies.

Laws that tackle disinformation often put pressure on the gatekeepers who exercise control over the infrastructure of the information environment. Many seek to define the responsibility of intermediaries, especially social media platforms, in curbing the spread of disinformation. In the EU, for example, the DSA creates several obligations for platforms to combat disinformation, such as promptly acting upon illegal content, preventing the dissemination of socially harmful content, and assessing the systemic risks of their services. It also creates obligations related to transparency and establishing communication with regulatory bodies and law enforcement agencies. The cooperation of intermediaries is vital for ensuring a coordinated response to disinformation, especially during critical times such as elections or public health emergencies.

When introducing new laws, there is a significant risk associated with adopting overly restrictive regulations, which could present a negative impact on human rights. The problem may be related to a lack of a clear definition of what disinformation is, as well as the concentration of power in the hands of the government. If public authorities have the power to define what information is 'false' in concrete cases, this may lead to censorship and undue restrictions on the rights to freedom of opinion and expression.

The Swedish approach - which does not seek to determine whether information is correct or incorrect, but places emphasis on the control over the information environment by foreign actors - offers a potential way to escape the pitfalls of defining what constitutes 'true' or 'false' information. Moreover, the existence of a body independent of the government, such as the PDA, can provide a viable institutional alternative to avoid government censorship. Nevertheless, this solution presupposes the resources to constantly monitor the information environment, in order to identify external threats and social vulnerabilities, as well as a distinction between external sources and domestic sources of (dis)information. This distinction is increasingly harder to achieve in the context of ever more sophisticated examples of astroturfing.

Governments may also enact laws that create the obligation to put in place media and information literacy initiatives, investing in the development of skills to combat disinformation. Some examples of binding obligations in this regard can be found in California, in the USA, and in the EU Audiovisual Media Services Directive (2018/1808), which requires member states to promote measures that develop media literacy skills.

Norms

Norms are social conventions that one often feels compelled to follow. They create societal expectations and influence behaviour. One of the ways to shape social norms is to put in place frameworks that serve as guidelines for action, even if non-binding. The European Charter for Media Literacy and the NAMLE Core Principles of Media Literacy Education are some examples.

MIL helps to shape social norms by fostering a culture of critical thinking and responsible information consumption. They are one of the most effective ways to embed anti-disinformation norms within society. Finland serves as an exemplary model in this regard, as the nation is ranked high in digital literacy skills. It is crucial for MIL initiatives to be dynamic and continuously updated to address new forms of disinformation and emerging technologies. For example, media literacy strategies should take into account the impact of AI-generated content on the disinformation landscape.

Beyond formal education, social norms are also reinforced through public awareness campaigns and civic engagement. Grassroots initiatives are an effective way to establish and reinforce norms. Community leaders and volunteers trained in media literacy can act as local information gatekeepers, helping to educate their communities, engaging in prebunking and debunking of disinformation. These initiatives not only spread important skills and knowledge but also build a sense of community responsibility and vigilance against disinformation. Lithuania provides an example of successful civic initiatives against disinformation.

Norms are further reinforced through the actions and policies of private sector entities, particularly social media platforms and tech companies. Platforms that implement transparent content moderation policies and provide users with tools and education to identify false information contribute to establishing societal expectations around responsible information consumption.

Media organisations and influencers also play a significant role in shaping norms. Responsible journalism and ethical reporting set standards for information quality and reliability. Media outlets that prioritise fact-checking and provide clear distinctions between news and opinion contribute to a culture of trust and accountability.

Market incentives

Financial incentives play an important role in the creation and dissemination of disinformation. There is a correlation between the economic incentives for disinformation and online advertising, and disinformation agents are further rewarded by the social media 'ad tech' industry.

On the one hand, platforms are motivated to combat disinformation not only to comply with regulations but also to maintain user trust and engagement. Users are less likely to engage with platforms that are perceived as unreliable or prone to spreading false information. Therefore, maintaining a clean and trustworthy information environment can be a competitive advantage for platforms. On the other hand, the revenue models of platforms based on advertising lead them to prioritise sensational content that drives user engagement, which can sometimes include disinformation.

So far, strategies to combat disinformation have not sufficiently taken into account the role of market incentives. The UN Global Principles for Information Integrity seek to change that, by mainstreaming the need to address the critical implications for information integrity resulting from business models that depend on targeted advertising based on behavioural tracking and personal data. It calls for a fundamental shift in incentive structures.

Infrastructure

Infrastructure, or the architecture of technological systems and tools, is essential for the rapid detection, mitigation, and prevention of disinformation. One of the most effective infrastructures in combating disinformation is the use of advanced technological tools, particularly those leveraging artificial intelligence and machine learning. Automated fact-checking systems and AI-driven content analysis can swiftly identify false information, preventing its widespread dissemination.

Platforms like Facebook and Twitter use AI to flag false content, while YouTube uses AI to monitor and flag harmful videos. However, over-reliance on AI without human oversight can lead to errors, especially in the context of less disseminated and spoken languages. These errors may lead to the removal of legitimate content, causing harm to freedom of expression.

A combination of AI and human judgment is essential. Another technological approach is watermarking AI-generated content, which embeds a unique identifier to trace the origin and authenticity of the material.

Effective infrastructure also requires global coordination and standardisation. Standardised protocols and best practices across platforms and countries can enhance the overall effectiveness of combating disinformation.

6.2. The need for a multistakeholder approach

A multistakeholder approach is akin to weaving a complex web, where each thread represents a different stakeholder. No single thread is sufficient on its own, but together they form a robust defence. When combined, these threads create a strong net capable of curbing disinformation. This collaborative effort is essential to building trust in the information ecosystem and empowering citizens to make informed decisions.

Governments

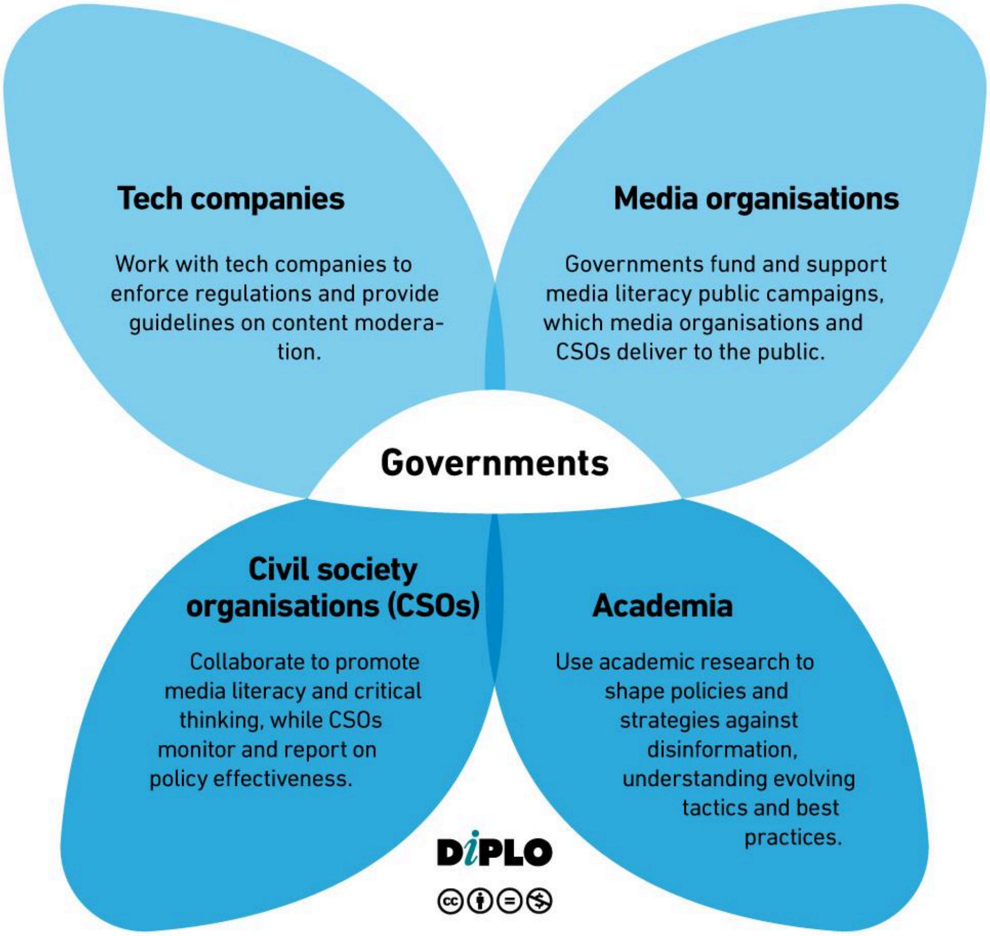
Governments lay the groundwork with laws, policies and guidelines that incentivise responsible online behaviour, especially from online platforms. While laws and regulations are important components in the fight against disinformation, they are not sufficient on their own.

First, mis- and disinformation evolve rapidly, with new methods and tactics constantly emerging. Laws and regulations, by their nature, can be slow to adapt. Second, mis- and disinformation are not confined by national borders. International cooperation and coordination are required to address the transnational nature of disinformation effectively. Governments may also act as conveners, bringing together diverse stakeholders to develop collaborative strategies. This is crucial given the complexity of disinformation, which spans national and international security, democracy, human rights, cybersecurity, and internet governance. Governments are able to facilitate dialogues among tech companies, media organisations, civil society organisations (CSOs), and academia to foster a unified approach. This involves creating forums for discussion, funding joint initiatives, and encouraging cross-sector partnerships.

Moreover, governments play a critical role in public education and awareness. They can include MIL in formal education curricula and launch nationwide campaigns to educate citizens about the risks of disinformation and how to recognise it. By promoting media literacy, governments empower individuals to critically assess the information they encounter, thus reducing the spread of false narratives.

Another significant aspect of the government's role is to support and protect independent fact-checking organisations. By providing resources and regulatory support, governments can ensure that these organisations have the capacity to rigorously debunk false information and provide the public with accurate data.

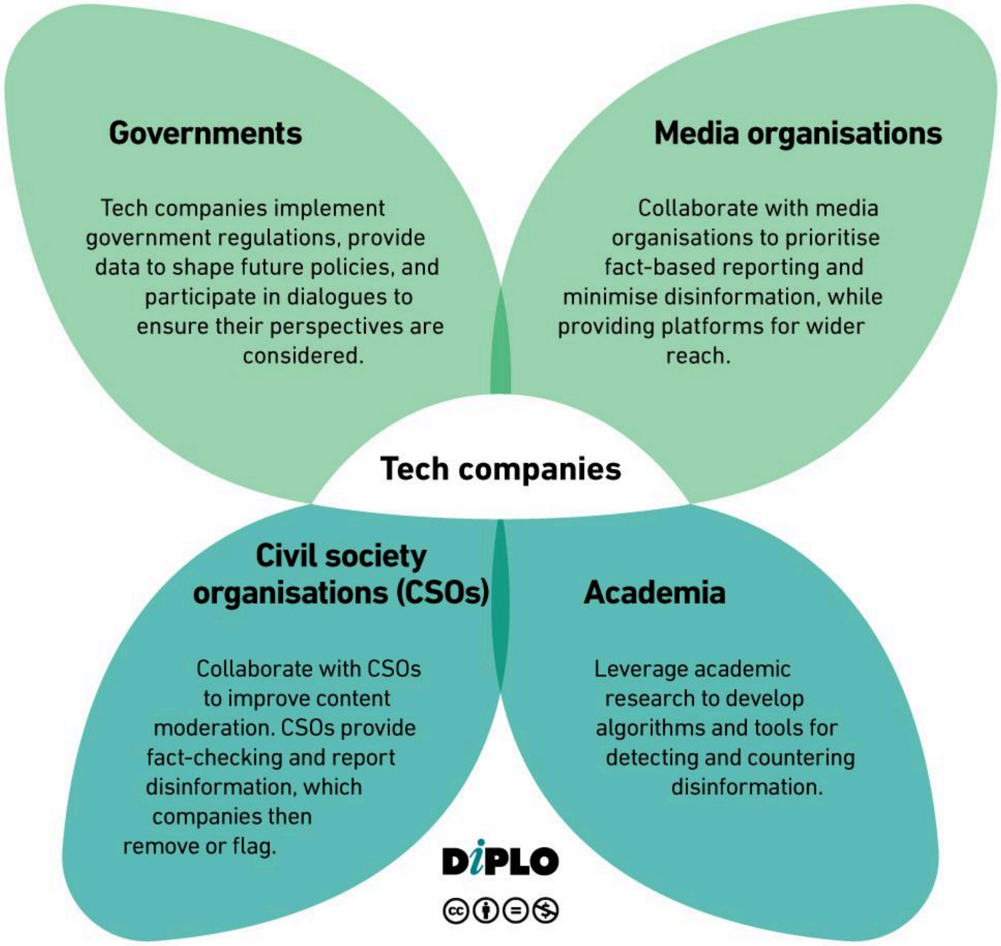
Governments also need to lead by example. This includes refraining from using disinformation laws as a way to censor speech. They could also invest in strengthening trust, not only by making sure that government information is accurate, timely, and easily accessible, but also by engaging in transparent policymaking processes and incorporating feedback from all relevant stakeholders to create balanced and effective disinformation strategies.


Tech companies
Work with tech companies to enforce regulations and provide guidelines on content moderation.
Media organisations
Governments fund and support media literacy public campaigns, which media organisations and CSOs deliver to the public.
Governments
Civil society organisations (CSOs)
Collaborate to promote media literacy and critical thinking, while CSOs monitor and report on policy effectiveness.
Academia
Use academic research to shape policies and strategies against disinformation, understanding evolving tactics and best practices.
DIPLO

Tech companies

Tech companies, particularly social media platforms, form the backbone of the information ecosystem. As such, these companies are not only well-placed to assist in combating disinformation that spreads through their infrastructure, but also in providing valuable data that could help researchers and policymakers better understand the phenomenon of disinformation, as well as in designing more effective policies to counter the issue.
Tech companies should continue to invest and perfect their content moderation tools, including those that employ AI, balancing automated solutions with the need for human review. Algorithms could be used to provide less visibility to content flagged as disinformation, whereas outright content removal should be reserved for the most serious cases.

It is also important for tech companies to collaborate with other stakeholders. Platforms can help media organisations reach broader audiences, and give priority to fact-based reporting. They could also collaborate with CSOs that provide fact-checking services and report disinformation. While respecting anonymity and data protection frameworks, companies can provide valuable data to researchers to study the spread and impact of disinformation. Increased transparency by tech companies and other information providers can enable a better understanding of how information is spread. The DSA is an example of a regulation that creates important data-sharing obligations.


Governments
Tech companies implement government regulations, provide data to shape future policies, and participate in dialogues to ensure their perspectives are considered.
Media organisations
Collaborate with media organisations to prioritise fact-based reporting and minimise disinformation, while providing platforms for wider reach.
Tech companies
Civil society organisations (CSOs)
Collaborate with CSOs to improve content moderation. CSOs provide fact-checking and report disinformation, which companies then remove or flag.
Academia
Leverage academic research to develop algorithms and tools for detecting and countering disinformation.
DiPLO

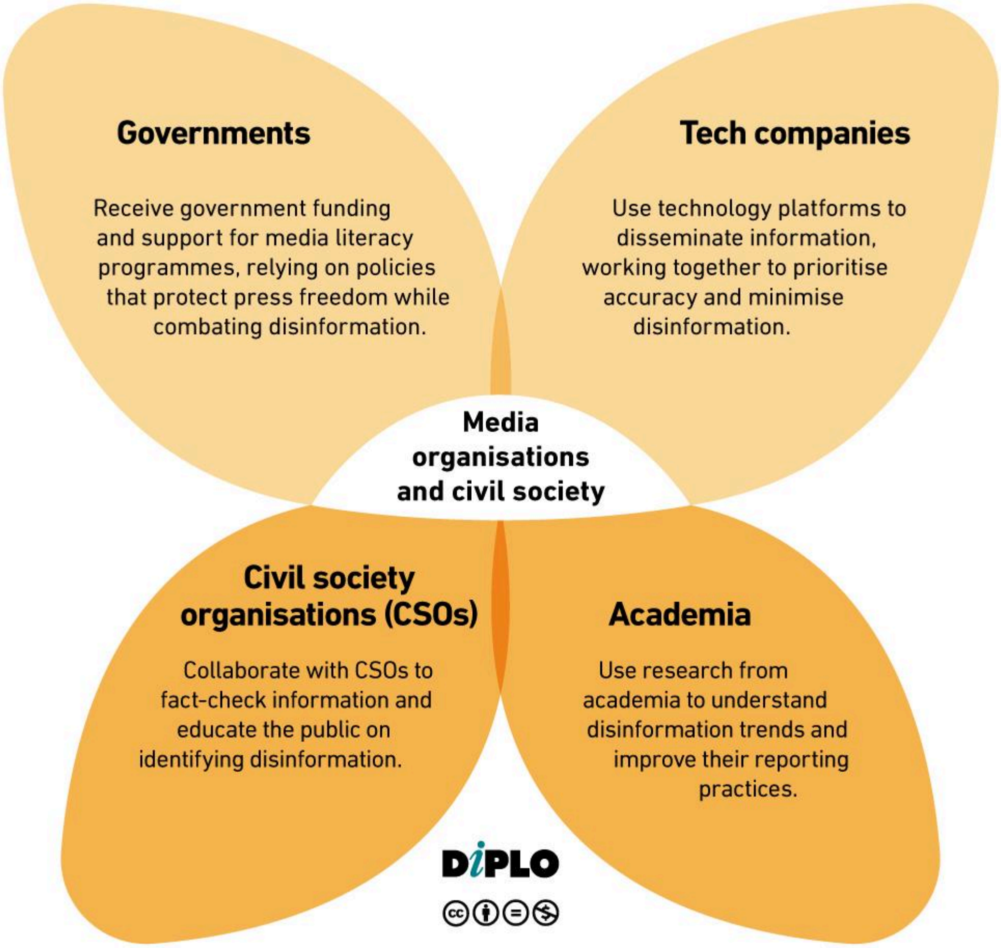
Media organisations and civil society

Media organisations and civil society play a crucial role in bringing other actors to account. Discussions on disinformation need to address the problem of state-sponsored disinformation, which can emanate from state institutions directly or from proxies targeting audiences within the state's own territory or abroad for political and strategic aims. When states systematically and simultaneously suppress other sources while promoting their own false narratives, they deny individuals the right to seek and receive information.

Media organisations and civil society can also help to place greater emphasis on the way that tech companies may undermine the right to freedom of opinion, manipulating the thought process necessary for someone to form his or her own opinion through the use of algorithms.

In addition, media organisations provide fact-based reporting and collaborate with fact-checkers to verify information. Accurate and ethical journalism is vital for maintaining the integrity of the information ecosystem. Media outlets must engage in fact-checking and collaborate with civil society to counter disinformation effectively, while governments must guarantee free press and media freedom. Public broadcasters can also play a role by educating viewers on how to identify and avoid disinformation.

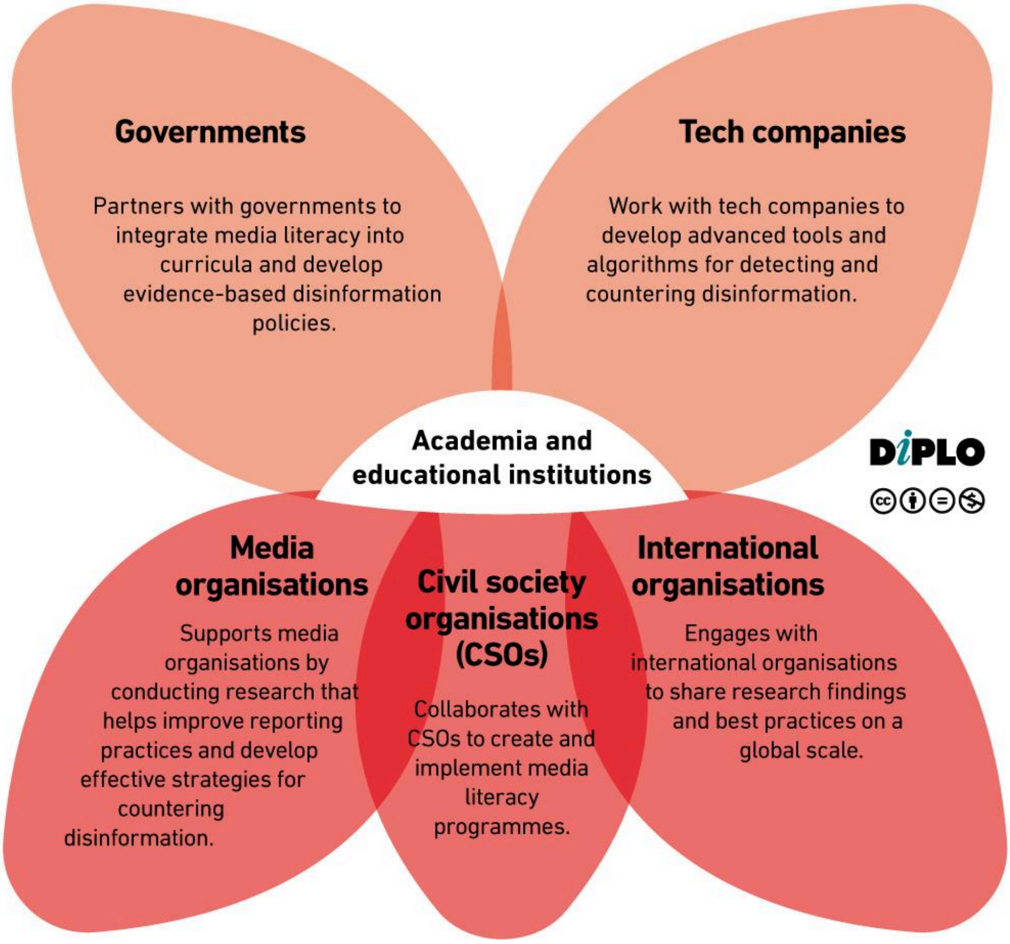
Civil society advocates for policies that support a healthy information environment. They play a vital role in fostering public awareness and conducting research on disinformation, bridging the gap between policymakers and the public, and ensuring that disinformation strategies are informed by grassroots insights and community needs (Wardle and Derakhshan, 2017).


Governments
Receive government funding and support for media literacy programmes, relying on policies that protect press freedom while combating disinformation.
Tech companies
Use technology platforms to disseminate information, working together to prioritise accuracy and minimise disinformation.
Media organisations and civil society
Civil society organisations (CSOs)
Collaborate with CSOs to fact-check information and educate the public on identifying disinformation.
Academia
Use research from academia to understand disinformation trends and improve their reporting practices.
DiPLO

Academia and educational institutions

Educational institutions are vital in fostering media literacy and critical thinking skills among the public, particularly the youth. Incorporating media literacy into the national school curriculum and also in lifelong learning makes individuals better equipped to identify and counteract disinformation.

Academic institutions contribute by developing educational materials that inform policy and public understanding. Research on the spread and impact of disinformation helps policymakers craft evidence-based strategies (Guess, Nyhan, & Reifler, 2020). The collaboration between Lund University in Sweden and the PDA is an emblematic example. Collaborative efforts between academia and government ensure that disinformation policies are adaptable and grounded in research.


Governments
Partners with governments to integrate media literacy into curricula and develop evidence-based disinformation policies.
Tech companies
Work with tech companies to develop advanced tools and algorithms for detecting and countering disinformation.
Academia and educational institutions
DiPLO
Media organisations
Supports media organisations by conducting research that helps improve reporting practices and develop effective strategies for countering disinformation.
Civil society organisations (CSOs)
Collaborates with CSOs to create and implement media literacy programmes.
International organisations
Engages with international organisations to share research findings and best practices on a global scale.

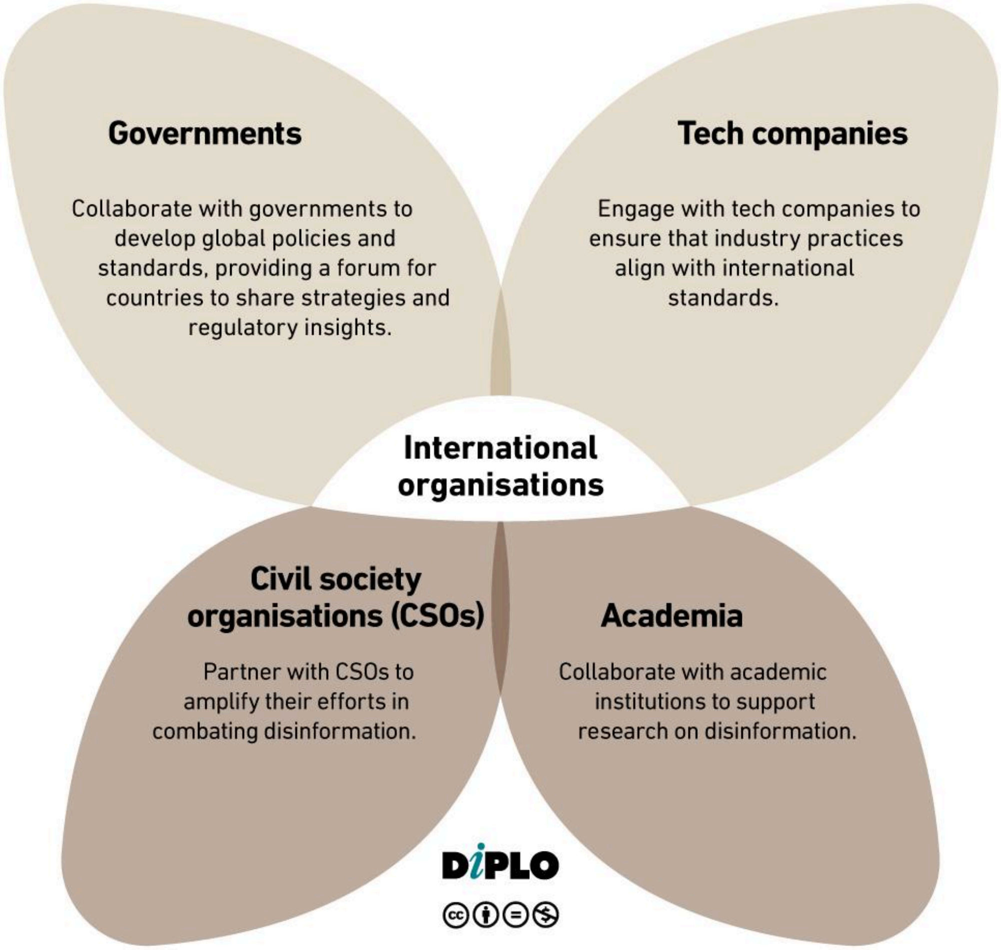
International organisations

One of the key roles of international organisations is to provide a platform for dialogue. They create forums where countries can share their experiences, strategies, and best practices, which fosters a collective understanding of the issue and helps develop effective solutions. By establishing global norms and standards, international organisations ensure that efforts to combat disinformation are aligned and mutually reinforcing. These guidelines and principles provide a common framework for action. Against this backdrop, the UN Global Principles for Information Integrity (United Nations, 2024) provides an important benchmark for the global harmonisation of priorities and basic approaches to disinformation.

Another important role of international organisations is to facilitate research and data sharing on disinformation trends and impacts. By aggregating and disseminating research findings, they help stakeholders understand the evolving nature of disinformation and the effectiveness of different countermeasures. This research-driven approach ensures that policies and strategies are evidence-based and can adapt to new challenges.

International organisations also emphasise inclusivity by advocating for the participation of diverse stakeholders, including smaller or less influential countries. This ensures that the solutions developed are equitable and effective across different contexts and regions. Furthermore, they mobilise resources and provide technical assistance to countries and stakeholders that may lack the capacity to combat disinformation effectively. This support includes funding, training, and expertise to enhance the implementation of national and regional initiatives.

International organisations also promote accountability and transparency among stakeholders. They monitor and evaluate the commitments and actions of governments, tech companies, and other actors, providing assessments and reports that highlight progress and areas for improvement. This oversight helps ensure that all stakeholders are held accountable for their role in combating disinformation and that their actions align with agreed-upon standards and principles.



Governments

Collaborate with governments to develop global policies and standards, providing a forum for countries to share strategies and regulatory insights.

Tech companies

Engage with tech companies to ensure that industry practices align with international standards.

International organisations

Civil society organisations (CSOs)

Partner with CSOs to amplify their efforts in combating disinformation.

Academia

Collaborate with academic institutions to support research on disinformation.

6.3. Concluding remarks: balancing the fight against disinformation with the protection of human rights and freedoms

Balancing the fight against disinformation with the protection of human rights is a complex and nuanced endeavour. Adhering to international human rights standards is the best way to combat disinformation while upholding the fundamental rights and freedoms that form the cornerstone of democratic societies.

Governments have a duty to protect human rights, including the rights to freedom of opinion and expression as enshrined in the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights. In order to observe human rights standards, governments must create an enabling environment for freedom of expression by promoting a free, independent, and diverse media landscape, which is essential for public debate and the open confrontation of ideas, as well as for combating disinformation.

The introduction of disinformation laws should seek to protect a legitimate and fundamental aim, and must be legal, proportionate, and necessary. Any limitation on freedom of expression must be exceptional and narrowly construed, and a higher threshold of legality, necessity, and proportionality should apply in the context of restrictions on political speech, including criticism of governments and political leaders, as well as other public figures. Improving platform transparency and accountability, along with providing media and digital literacy, could sometimes be a better course of action than enacting specific regulations on disinformation, especially if disinformation is vaguely defined.

Information campaigns that target ethnic minorities, immigrants, and other marginalised communities, exacerbating social tensions and leading to hate crimes, are to be prohibited by law, regardless of any assessment of truthfulness, as per Article 20 (2) of the ICCPR. Public authorities and companies alike are under the obligation to act against such content.

Concerns about disinformation often relate to influence operations originating from abroad. Although this is an important aspect, more attention could be paid to government policies that negatively affect populations within the country. This is the case with measures to combat disinformation that go overboard, or government-sponsored disinformation. In different ways, both may impinge on freedom of opinion and expression.

Problems related to information operations are increasingly being framed as a matter of national security. The most important consequence of identifying a problem as a security issue is that it is given priority and urgency in relation to other issues on the public agenda, requiring emergency measures and justifying actions outside the normal bounds of political procedure (Buzan et al. 1998). As a collateral effect, there could be a decrease in accountability and public control over decisions made to fight disinformation. In order to counter this trend, governments should engage in transparent policymaking processes, incorporating feedback from all relevant stakeholders to create balanced and effective disinformation strategies.

While governments have an obligation to protect human rights, companies have a responsibility to *respect* human rights standards. Companies are expected to conduct human rights risk assessments and due diligence, ensuring their business models and operations do not negatively impact human rights. This includes the sharing of data and information on algorithms, which could make an assessment of the correlation between the spread of disinformation and ‘ad tech’ business models possible. Regulation has been focused on the responsibility of intermediaries to curb the spread of mis- and disinformation, but more needs to be done to change the incentive structures that underpin the business model of the ‘ad tech’ industry and social media platforms.

Companies should adopt transparent content moderation practices and manage algorithmic systems responsibly to maintain the integrity of information on their platforms. One of the significant challenges for companies is balancing content moderation with freedom of expression. Overly aggressive content removal policies can lead to the suppression of legitimate speech. Therefore, companies must ensure that their moderation practices are transparent, consistent, and based on clear guidelines that respect human rights. The EU Strengthened Code of Practice on Disinformation provides valuable suggestions on how platforms could assist in combating disinformation while respecting human rights standards.

Civil society and journalists also play an important role in this landscape. Civil society organisations, journalists, and human rights defenders are crucial in holding both governments and companies accountable. They provide independent oversight, contribute to public education on media literacy, and work to debunk mis- and disinformation. CSOs often spearhead fact-checking initiatives and promote media literacy programmes that equip individuals with the skills to critically evaluate information. They also advocate for stronger protections of freedom of expression and the establishment of environments in which diverse opinions can be freely expressed and debated.

Policy approaches to combating mis- and disinformation have been based on promoting participation and agency through media literacy strategies, on the one hand, and on protecting society through regulation and content policy, on the other. These two approaches should be seen as complementary and mutually reinforcing. At the individual level, media literacy strengthens resilience against false information, focusing on knowledge acquisition and attitude transformation. It empowers individuals to engage in prebunking and debunking and to become active participants in combating misinformation. At the collective level, content norms aim to strengthen societal resistance by acknowledging the need to provide safeguards against information disorder. Media literacy is a cornerstone of societal trust and resilience, of the principles enshrined in the UN Global Principles for Information Integrity (2024).

Striking the right balance between protection and participation in combating disinformation means resorting wisely to both regulation and engagement. The latter should be conceived in broad terms, encompassing not only the active involvement of individuals, but also the involvement of other segments such as educators, companies, and technical actors. This inclusive approach provides a pathway to curb disinformation while respecting human rights.

7. References

Academia Singapore (2019) POFMA: Letter to Education Minister. 11 April. Available at https://www.academia.sg/pofma-letter/

Ahmed T (2023) Minnesota advances deepfakes bill to criminalize people sharing altered sexual, political content. Associated Press, 11 May. Available at https://apnews.com/article/deepfake-minnesota-pornography-elections-technology-5ef76fc3994b2e437c7595c09a38e848

Allyn B (2021) Stung By Twitter, Trump Signs Executive Order To Weaken Social Media Companies. *Npr*, 28 May. Available at https://www.npr.org/2020/05/28/863932758/stung-by-twitter-trump-signs-executive-order-to-weaken-social-media-companies

Amnesty International (2023) Singapore: Suppression of activists, critics continue ahead of elections. Public Statement, 16 May. Available at https://www.amnesty.org/ar/wp-content/uploads/2023/05/ASA3667882023ENGLISH.pdf

ASEAN (2022) Training of Trainers Program to Address Disinformation and Promote Media Literacy: Toolkit for Educators. Jakarta: ASEAN Secretariat. Available at https://asean.org/wp-content/uploads/2022/01/TOT-Program-to-Address-Disinformation_with-ISBN.pdf

ASEAN (2024) ASEAN Guide on AI Governance and Ethics. Available at https://asean.org/book/asean-guide-on-ai-governance-and-ethics/

ASEAN Foundation (2022) ASEAN Digital Literacy Programme. 11 February. Available at https://www.aseanfoundation.org/asean_digital_literacy_programme

ASEAN Ministers Responsible for Information (2014) Declaration on Social Responsible Media for a Peaceful and Prosperous ASEAN Community. Available at https://asean.org/wp-content/uploads/2021/01/Declaration-on-Social-Responsible-Media-for-a-Peaceful-and-Prosperous-ASEAN-Community.pdf

ASEAN Ministers Responsible for Information (2018) Framework and Joint Declaration to Minimise the Harmful Effects of Fake News. Available at https://asean.org/wp-content/uploads/2021/01/Framework-and-Joint-Declaration-to-Minimise-the-Harmful-Effects-of-Fake-News.pdf

ASEAN Ministers Responsible for Information (2018) Core Values on Digital Literacy for ASEAN. Available at https://asean.org/wp-content/uploads/2022/03/CORE_VALUES_ON_DIGITAL_LITERACY_FOR_ASEAN.pdf

Association for Progressive Communications APC (2021) Disinformation and freedom of expression. Available at https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/APC-Disinformation-Submission.pdf

Barlow J P (1996) A Declaration of the Independence of Cyberspace. Available at https://www.eff.org/cyberspace-independence

Bjurling B et al. (2024) Foreign Information Manipulation & Interference A Large Language Model Perspective. RISE Research Institutes of Sweden AB. Available at https://mpf.se/psychological-defence-agency/publications/archive/2024-03-15-foreign-information-manipulation--interference-a-large-language-model-perspective

Bradshaw S and Neudert L (2021) *The Road Ahead: Mapping Civil Society Responses to Disinformation*. National Endowment for Democracy. Available at https://www.ned.org/mapping-civil-society-responses-to-disinformation-international-forum/.

Brannon V C and Holmes E R (2024) Section 230: An Overview. Congressional Research Service R46751. Available at https://crsreports.congress.gov/product/pdf/R/R46751

Braun K (2019) Unpacking post-truth. *Critical Policy Studies Vol 13 (4)*. Available at https://www.tandfonline.com/doi/full/10.1080/19460171.2019.1673200?scroll=top&needAccess=true

Broadband Commission (2020) *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression*. Available at https://www.broadbandcommission.org/Documents/working-groups/ExecSum_FoE_disinfo_report.pdf

Budak C et al. (2024) Misunderstanding the harms of online misinformation. *Nature*, Vol 630 no. 6.

Buzan, B et al. *Security: A New Framework for Analysis* London: Lynne Rienner Publishers, 1998.

Cappelletti J and Svenson A (2023) Michigan to join state-level effort to regulate AI political ads as federal legislation is pending. Associated Press, 29 November. Available at https://apnews.com/article/michigan-political-ads-artificial-intelligence-94fdb689b8c44b1f2818b57975bbfa9c

Center for Countering Digital Hate [CCDH] (2024) Attack of the Voice Clones. Available at https://counterhate.com/wp-content/uploads/2024/05/240524-Attack-of-the-Voice-Clones-REPORT_final.pdf

Center for Democracy and Technology [CDT] (2010) Intermediary Liability: Protecting Internet Platforms for Expression and Innovation. Available at https://cdt.org/wp-content/uploads/pdfs/CDT-Intermediary%20Liability_(2010).pdf

Chambers S (2021) Truth, Deliberative Democracy and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere? *Political Studies Volume 69 Issue 1.* Available at https://journals.sagepub.com/doi/10.1177/0032321719890811

Chander A and Sun H (Eds.) (2023*). From the digital Silk Road to the return of the state*. New York: Oxford University Press.

Cheng JH and Chow M (2023) Strengthening Cyber Resilience in Southeast Asia. *Fulcrum*, 6 November. Available at https://fulcrum.sg/strengthening-cyber-resilience-in-southeast-asia/#:~:text=Singapore%20and%20Malaysia%20rank%20first,Act%2C%20and%20Computer%20Misuse%20Act.

Cheong (2023) Singapore has recognized the real danger of disinformation. *Nikkei Asia*, 9 November. Available at https://asia.nikkei.com/Opinion/Singapore-has-recognized-the-real-danger-of-disinformation

Civic Resilience Initiative et al. (2021) Resilience to Disinformation. The Lublin Triangle Perspective. Available at https://mfa.gov.ua/storage/app/sites/1/Docs/the-lublin-triangle-joint-report-on-countering-disinformation.pdf

Civic Resilience Initiative et al, (2021) Challenges of Contemporary Disinformation. Available at https://cri.lt/wp-content/uploads/2019/10/Challenges-EN.pdf

Collaboration on International ICT Policy in East and Southern Africa [CIPESA] (2019) Desposts and Disruptions: Five Dimensions of Internet Shutdowns in Africa. Available at https://cipesa.org/wp-content/files/briefs/report/Despots-And-Disruptions_March-20.pdf

Colomina C et al. (2021) The impact of disinformation on democratic processes and human rights in the world. European Parliament. Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf

Conger K (2023) Twitter to Relax Ban on Political Ads. *The New York Times,* 3 January. Available at https://www.nytimes.com/2023/01/03/technology/twitter-political-ads.html

Cornish P (Ed) (2021) The Oxford Handbook of Cybersecurity. Oxford: Oxford University Press.

Coulter M (2024) Google to launch anti-misinformation campaign ahead of EU elections. *Reuters*, 16 February. Available at https://www.reuters.com/technology/google-launch-anti-misinformation-campaign-ahead-eu-elections-2024-02-16/

Dang S and Paul K (2024) OpenAI, Meta and other tech giants sign effort to fight AI election interference. *Reuters*, 16 February. Available at https://www.reuters.com/technology/openai-meta-other-tech-giants-sign-effort-fight-ai-election-interference-2024-02-16/

DeJournette T (2024) Hawaii Legislators Seek To Crack Down On AI-Generated Political Misinformation. *Honolulu Civil Beat*, 19 February. Available at https://www.civilbeat.org/2024/02/hawaii-legislators-seek-to-crack-down-on-ai-generated-political-misinformation/

DeNardis L and Musiani F (2016) Governance by Infrastructure. In: *Musiani F et al. (eds.) The Turn to the Infrastructure in Internet Governance*. New York: Palgrave Macmillan.

Digwatch (2022) EU's Digital Services Act (DSA) & Digital Markets Act (DMA). Available at https://dig.watch/processes/eu-digital-service-act-dsa-digital-market-act-dma

European Commission (2018a) Action Plan against Disinformation. Brussels, JOIN(2018) 36 final. Available at https://www.eeas.europa.eu/sites/default/files/action_plan_against_disinformation.pdf

European Commission (2020) Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions On the European democracy action plan. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A790%3AFIN&qid=1607079662423

European Commission (2022a) Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training. Directorate-General for Education, Youth, Sport and Culture. Publications Office of the European Union. Available at https://data.europa.eu/doi/10.2766/28248

European Commission (2022b) Strengthened Code of Practice on Disinformation. Available at https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

European Commission (2024a), Commission sends requests for information on generative AI risks to 6 Very Large Online Platforms and 2 Very Large Online Search Engines under the Digital Services Act. Available at https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-generative-ai-risks-6-very-large-online-platforms-and-2-very

European Commission (2024b) Commission publishes guidelines under the DSA for the mitigation of systemic risks online for elections. Press Release, 26 March. Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707

European Digital Media Observatory (EDMO) Finland. Available at https://edmo.eu/resources/repositories/mapping-the-media-literacy-sector/finland/

European External Action Services [EEAS] (2023) Disinformation: Opening speech by High Representative/Vice-President Josep Borrell at the EEAS Conference on Foreign Information Manipulation and Interference. 7 February. Available at https://www.eeas.europa.eu/eeas/disinformation-opening-speech-high-representativevice-president-josep-borrell-eeas-conference_en

European Parliament (2024) Parliament adopts new transparency rules for political advertising *News European Parliament*, 27 February. Available at https://www.europarl.europa.eu/news/en/press-room/20240223IPR18071/parliament-adopts-new-transparency-rules-for-political-advertising

European Union (2022) Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Available at https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Union (2024) Regulation (Eu) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Evenett S and Fritz J (2022) Emergent Digital Fragmentation: the perils of unilateralism. Global Trade Alert. Available at https://digitalpolicyalert.org/report/emergent-digital-fragmentation

Falkheimer J et al. (2023) Malign foreign interference and information influence on video game platforms: Understanding the adversarial playbook. Psychological Defence Agency and Lund University. Available at https://mpf.se/psychological-defence-agency/publications/archive/2023-12-01-malign-foreign-interference-and-information-influence-on-video-game-platforms-understanding-the-adversarial-playbook

Fedorov M (2022) Letter from the Deputy Prime Minister of Ukraine to Goran Marby, President and Chief Executive Officer of ICANN. Available at https://www.icann.org/en/system/files/correspondence/fedorov-to-marby-28feb22-en.pdf

Fee J (2021) Resilience Against the Dark Arts: A Comparative Study of British and Swedish Government Strategies Combatting Disinformation. Available at https://www.diva-portal.org/smash/record.jsf?dswid=-8557&pid=diva2%3A1564473

Finnish Ministry of Education and Culture (2019) Media Literacy in Finland. National Media Education Policy. Available at https://medialukutaitosuomessa.fi/mediaeducationpolicy.pdf

Finnish Ministry of Foreign Affairs (2024) Finland and United States sign a Memorandum of Understanding on countering foreign state information manipulation. Finnish Government, 4 April. Available at https://valtioneuvosto.fi/en/-/finland-and-united-states-sign-a-memorandum-of-understanding-on-countering-foreign-state-information-manipulation

Fitriani et al. (2024) Policy Brief: Regional and Cross-Border Responses Towards Disinformation in Southeast Asia, Safer Internet Lab (SAIL). Available at https://saferinternetlab.org/wp-content/uploads/2024/04/Policy-Brief-Regional-and-Cross-Border-Responses-Towards-Disinformation-in-Southeast-Asia-.pdf

Frau-Meigs D and Corbu N (2024) *Disinformation Debunked: Building Resilience through Media and Information Literacy*. Abingdon: Routledge.

Friggeri A et al. (2014) Rumor Cascades. In: *Proceedings of the international AAAI conference on web and social media*. Available at https://ojs.aaai.org/index.php/ICWSM/article/view/14559

Freedom House (2023) Singapore in Freedom on the Net 2023. Available at https://freedomhouse.org/country/singapore/freedom-net/2023

Freedom House (2024) Singapore in the Freedom in the World 2024. Available at https://freedomhouse.org/country/singapore/freedom-world/2024

Gagliardone I and Stremlau N (2022) It’s Time to Revisit the Framing of Internet Shutdowns in Africa. Carnegie Endowment for International Peace, 21 November. Available at https://carnegieendowment.org/posts/2022/11/its-time-to-revisit-the-framing-of-internet-shutdowns-in-africa?lang=en

Glenski M et al. (2018) Propagation From Deceptive News Sources Who Shares, How Much, How Evenly, and How Quickly?, in *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp. 1071-1082.

Global Disinformation Index [GDI] (2020) Research Brief: Ad Tech Fuels Disinformation Sites in Europe – The Numbers and Players. Available at https://www.disinformationindex.org/research/2020-3-1-research-brief-ad-tech-fuels-disinformation-sites-in-europe-the-numbers-and-players/

Government Offices of Sweden, Ministry of Defence (2021) Main elements of the Government bill Totalförsvaret 2021–2025 Total defence 2021–2025. Available at https://www.swedenabroad.se/globalassets/ambassader/nederlanderna-haag/documents/government-bill-totalforsvaret-20212025.pdf

Government Offices of Sweden, Prime Minister's Office (2024) National Security Strategy. Available at
https://www.government.se/globalassets/government/national-security-strategy.pdf

Grabe M A. and Bucy E (2023) Towards diagnostics and mitigation of polluted information ecosystems: Historical and philosophical considerations. Available at
https://www.researchgate.net/publication/374022512_Towards_diagnostics_and_mitigation_of_polluted_information_ecosystems_Historical_and_philosophical_considerations

Griera M (2024) EU parties and Commission sign campaign rulebook against foreign interference, disinformation. Euractiv, 9 April. Available at
https://www.euractiv.com/section/elections/news/eu-parties-and-commission-sign-campaign-rulebook-against-foreign-interference-disinformation/

Grizzle A et al. (2021) Media and information literate citizens: think critically, click wisely! Paris: UNESCO. Available at https://unesdoc.unesco.org/ark:/48223/pf0000377068

Healy J (2021) These Are the 5 People Who Died in the Capitol Riot. *New York Times,* January 11. Available at
https://www.nytimes.com/2021/01/11/us/who-died-in-capitol-building-attack.html

Hameleers M (2024) The state of the art in combating mis- and disinformation: lessons from pre- and debunking approaches. In: Frau-Meigs D and Corbu N. *Disinformation Debunked: Building Resilience through Media and Information Literacy*. Abingdon: Routledge.

Herasimenka A et al. (2023). The political economy of digital profiteering: communication resource mobilization by anti-vaccination actors. *Journal of Communication*. Available at
https://academic.oup.com/joc/article/73/2/126/6960639

Hjort J and Tian L (2024) The Economic Impact of Internet Connectivity in Developing Countries. *PEDL Synthesis Series n. 6.* Available at
https://pedl.cepr.org/sites/default/files/Synthesis%20Paper%20SP6%20Jonas%20Hjort.pdf

Hsu T et al. (2024) Elections and Disinformation Are Colliding Like Never Before in 2024. *The New York Times*, 9 January. Available at
https://www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html

Hyvärinen J (2024) Hostile Information Campaigns Could Test a Divided Finland. *Tech Policy.Press*, 30 May. Available at
athttps://www.techpolicy.press/hostile-information-campaigns-could-test-a-divided-finland/

Indonesian Ministry of Communication and Informatics (2023) ASEAN Guideline on Management of Government Information in Combating Fake News and Disinformation in the Media. Available at
https://asean.org/book/asean-guideline-on-management-of-government-information-in-combating-fake-news-and-disinformation-in-the-media/

Infocomm Media Development Authority Singapore (2023) IMDA's Online Safety Code comes into effect, 17 July. Available at
https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/imdas-online-safety-code-comes-into-effect

Instagram (2024) Continuing our Approach to Political Content on Instagram and Threads. *Blog Post*, 9 February. Available at https://about.instagram.com/blog/announcements/continuing-our-approach-to-political-content-on-instagram-and-threads

International Commission of Jurists (2021) Dictating the Internet: a Human Rights Assessment of the Implementation of Singapore's Protection from Online Falsehoods and Manipulation Act 2019. Available at /https://icj2.wpenginepowered.com/wp-content/uploads/2021/10/Singapore-Dictating-the-Internet-Legal-Briefing-2021-ENG.pdf

Khalid A (2019) Americans can’t stop relying on social media for their news. *Quartz*, 15 May. Available at https://qz.com/1720695/pew-study-shows-more-americans-rely-on-social-media-for-news

Kivinen L (2022) Pragmatism Defeats Propaganda – Finland’s Move to NATO. Center for European Policy Analysis (CEPA). 20 May. Available at https://cepa.org/article/pragmatism-defeats-propaganda-finlands-move-to-nato/

Kurbalija J (2004) The Classification of Internet Governance, DiploFoundation. Available at https://www.diplomacy.edu/wp-content/uploads/2014/01/Internet_Governance_Classification_ver_07102004.pdf

Langguth J et al. (2023) COVID-19 and 5G conspiracy theories: long term observation of a digital wildfire. *International Journal of Data Sciences and Analytics*, 15(3). Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9137448/

Lee Myers S (2023) Sweden Is Not Staying Neutral in Russia’s Information War. *The Washington Post*, 10 August. Available at https://www.nytimes.com/2023/08/10/technology/sweden-combat-disinformation.html

Lessenski M (2018) Common Sense Wanted: resilience to ‘post-truth’ and its predictors In: The New Media Literacy Index 2018. Open Society Institute, Sofia. Available at https://osis.bg/wp-content/uploads/2018/04/MediaLiteracyIndex2018_publishENG.pdf

Lessig L (2006) *Code: version 2.0*. New York: Basic Books.

Levine D and Savoia L (2023) Slovakia’s election deep fakes show how AI could be a danger to U.S. elections. *Fulcrum*, 28 November. Available at https://thefulcrum.us/slovakias-election-deep-fakes-show-how-ai-could-be-a-danger-to-us-elections

Loomba S (2021) Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour,* Vol. 5, March. Available at https://www.nature.com/articles/s41562-021-01056-1

Luhmann N (1992) What is communication? *Communication Theory* 2 3

Jones-Jang M et al. (2019) Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don’t. *American Behavioral Scientist* 1–18. Available at https://www.researchgate.net/publication/335352499_Does_Media_Literacy_Help_Identification_of_Fake_News_Information_Literacy_Helps_but_Other_Literacies_Don't

Kozłowski A (2024) Combatting disinformation by state agencies: the case of the Swedish Psychological Defence Agency. *New Eastern Europe,* 7 May. Available at https://neweasterneurope.eu/2024/05/07/combatting-disinformation-by-state-agencies-the-case-of-the-swedish-psychological-defence-agency/

Mackey A (2021) Newly Released Records Show How Trump Tried to Retaliate Against Social Media For Fact-Checking. *Electronic Frontier Foundation*, 28 May. Available at https://www.eff.org/deeplinks/2021/05/newly-released-records-show-how-trump-tried-retaliate-against-social-media-fact

Maréchal N et al. (2020) Getting to the Source of Infodemics: It's the Business Model. New America. Available at https://d1y8sb8igg2f8e.cloudfront.net/documents/Getting_to_the_Source_of_Infodemics_Its_the_Business_Model_2020-05.pdf

Mays B (2023) The Disinformation Landscape in Lithuania. *EU Disinfo Lab*. Available at https://www.disinfo.eu/wp-content/uploads/2023/06/20230521_LT_DisinfoFS.pdf

Meaker M (2024) Slovakia's Election Deepfakes Show AI Is a Danger to Democracy. *Wired*, 3 October. Available at https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/

Moncau L (2021) Direito ao Esquecimento: Entre a liberdade de expressão, a privacidade e a proteção de dados pessoais. São Paulo: Editora Revista dos Tribunais.

Morgan K (2024) Our work to prepare for the 2024 European elections. *TikTok Blog Post*, 14 February. Available at https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections

Open Society Institute Sofia (2022). How It Started, How It is Going: Media Literacy Index 2022. Policy Brief 57 - October. Available at https://osis.bg/wp-content/uploads/2022/10/HowItStarted_MediaLiteracyIndex2022_ENG_.pdf

Organization for Economic Co-operation and Development [OECD] (2023) Media Literacy Education System: Finland. The OECD Dis/Mis Resource Hub. Available at https://www.oecd.org/en/publications/2023/02/mis-and-disinformation_c584ff91/media-literacy-education-system_28535133.html

Palmertz B et al. (2024) Building Resilience and Psychological Defence An analytical framework for countering hybrid threats and foreign influence and interference. Lund University Psychological Defence Research Institute Working Paper 2024:1. Available at https://www.psychologicaldefence.lu.se/sites/psychologicaldefence.lu.se/files/2024-03/BuidlingResilienceAndPsychologicalDefence.pdf

Parliamentary Assembly of the Council of Europe [PACE] (2020) Democracy hacked? How to respond? Resolution 2326 31 January. Available at https://pace.coe.int/en/files/28598/html

Pawelec M (2022) Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions. *Springer*. Available at https://link.springer.com/article/10.1007/s44206-022-00010-6

Pettersson H (2018) Media Literacy Across Europe In: Mackintosh E (2018) Finland is winning the war on fake news. What it's learned may be crucial to Western democracy. *CNN*. Available at https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/

Pew Research (2022) Social Media Seen as Mostly Good for Democracy Across Many Nations, But U.S. is a Major Outlier. Available at https://www.pewresearch.org/wp-content/uploads/sites/20/2022/12/PG_2022.12.06_Online-Civic-Engagement_REPORT.pdf

Phartiyal S (2019) Social media fake news fans tension between India and Pakistan. *Reuters*, 28 February. Available at https://www.reuters.com/article/idUSKCN1QH1N4/

Polish Ministry of Foreign Affairs (2021) Declaration of the Lublin Triangle Foreign Ministers of joint European heritage and common values. 7 July. Available at https://www.gov.pl/web/diplomacy/declaration-of-the-lublin-triangle-foreign-ministers-of-joint-european-heritage-and-common-values

Psychological Defence Agency [PDA] (undated) About us. Available at https://mpf.se/psychological-defence-agency/about-us/our-mission

Psychological Defence Agency (2024) Countering information influence activities. Available at https://mpf.se/psychological-defence-agency/publications/archive/2024-03-01-countering-information-influence-activities

Ray A (2021) Disinformation, Deepfakes and Democracies: The Need for Legislative Reform. *UNSW Law Journal* 44(3).Available at https://www.unswlawjournal.unsw.edu.au/wp-content/uploads/2021/09/Issue-443_final_Ray.pdf

Reporters Without Borders [RSF] (2019) RSF explains why Singapore's anti-fake news bill is terrible. Available at https://rsf.org/en/rsf-explains-why-singapore-s-anti-fake-news-bill-terrible

Roozenbeek J and van der Linden S (2019) Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 65. Available at https://www.nature.com/articles/s41599-019-0279-9

Robins-Early N (2024) Google restricts AI chatbot Gemini from answering questions on 2024 elections. *The Guardian*, 12 March. Available at https://www.theguardian.com/us-news/2024/mar/12/google-ai-gemini-2024-election

Rosenthal HM (2023) Speech Imperialization? Situating American Parrhesia in an Isegoria *World. International Journal for the Semiotics of Law*, 35:583–603

Rothkopf DJ (2003) When the buzz bites back. *The Washington Post*, 11 May. Available at http://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html

Singapore Legal Advice (2022) Singapore Fake News Laws: Guide to POFMA (Protection from Online Falsehoods and Manipulation Act). *Singapore Legal Advice.com*, 25 April. Available at https://singaporelegaladvice.com/law-articles/singapore-fake-news-protection-online-falsehoods-manipulation/#:~:text=POFMA%20prohibits%20the%20communication%20of,be%20a%20representation%20of%20fact.

Singapore Ministry of Communications and Information (2018), Digital Media and Information Literacy Framework. Available at https://www.mci.gov.sg/files/mci%20dmil%20framework.pdf

Singapore Ministry of Communications and Information (2023), Online Safety (Miscellaneous Amendments) Act Takes Effect on 1 February 2023. Available at https://www.mci.gov.sg/media-centre/press-releases/online-safety-act-takes-effect-on-1-february-2023/

Singapore Ministry of Education (2024), Applied Learning Programme. Available at https://www.moe.gov.sg/secondary/courses/express/electives/?term=Applied%20Learning%20Programme%20(ALP)

Singapore Ministry of Education (2024), Artificial Intelligence. Available at https://www.moe.gov.sg/news/parliamentary-replies/20230109-artificial-intelligence

Singapore Ministry of Education (2024), Practicing Cyber Wellness. Available at https://www.moe.gov.sg/education-in-sg/our-programmes/cyber-wellness

Singapore Ministry of Home Affairs (2023) Disinformation and Influence Campaigns. Available at https://www.mha.gov.sg/docs/default-source/default-document-library/03_22-ne-insights-on-disinformation-and-influence-campaigns.pdf?sfvrsn=24e85b5c_2

Singapore Ministry of Law (2019) New Bill to Protect Society from Online. Falsehoods and Malicious Actors. Available at https://www.mlaw.gov.sg/news/press-releases/new-bill-to-protect-society-from-online-falsehoods-and-malicious-actors/

Singh M (2020) Russia and Iran obtained US voter data in bid to sow unrest before election, FBI warns. *The Guardian*, 22 October. Available at https://www.theguardian.com/us-news/2020/oct/21/russia-iran-us-voter-data-security-fbi-election

Smith-Spark L and Vandoorne S (2018) As France debates fake news bill, critics see threat to free speech. *CNN*. Available at https://edition.cnn.com/2018/06/07/europe/france-fake-news-law-intl/index.html

Sörensen S and Pamment J (2023) Operationalising the Framework for Evaluating Capability Against Information Influence Operations A Case Study of the Psychological Defence Agency’s Courses. NATO Strategic Communications Centre Of Excellence. Available at https://stratcomcoe.org/publications/operationalising-the-framework-for-evaluating-capability-against-information-influence-operations-a-case-study-of-the-psychological-defence-agencys-courses/295

Statista (2024) Share of adults who use social media as a source of news in selected countries worldwide as of February 2024. Available at https://www.statista.com/statistics/718019/social-media-news-source/

Timm C (2022) Strategies and Resilience Against Disinformation Influences How Germany and Sweden respond to Russian disinformation. Lund University. Available at https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9070305&fileOId=9072742

Think Centre (TC) et.al (2011) Universal Periodic Review on Singapore for the 11th Session of UPR. Available at https://www.ohchr.org/sites/default/files/lib-docs/HRBodies/UPR/Documents/session11/SG/JS1_JointSubmission1-eng.pdf

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2013) Global Media and Information Literacy Assessment Framework: Country Readiness and Competencies. Paris: UNESCO. Available at https://allchildrenlearning.org/wp-content/uploads/2019/12/Global-Media-and-Information-Literacy-Assessment-Framework_-country-readiness-and-competencies-UNESCO-Digital-Library.pdf

Ukrainian Ministry of Foreign Affairs (2020). Joint Declaration of Foreign Ministers of the Republic of Poland, the Republic of Lithuania and Ukraine on establishing Lublin Triangle. 28 July. Available at https://mfa.gov.ua/en/news/joint-declaration-foreign-ministers-republic-poland-republic-lithuania-and-ukraine-establishing-lublin-triangle

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2018) Journalism, ‘Fake News’ & Disinformation. Handbook for Journalism Education and Training. Paris: UNESCO. Available at https://unesdoc.unesco.org/ark:/48223/pf0000265552

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2019) Journalism and Elections in Times of Disinformation. Addis Ababa Declaration World Press Freedom Day 2019. Available at https://au.int/sites/default/files/pressreleases/36586-pr-wpfdaddisdecl3_may.pdf

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2022a) UNESCO works to counter mis- and disinformation. Available at https://unesdoc.unesco.org/ark:/48223/pf0000382385

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2022b) Elections in digital times: a guide for electoral practitioners. Available at https://unesdoc.unesco.org/ark:/48223/pf0000382102

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2023) Online disinformation: UNESCO unveils action plan to regulate social media platforms. Available at https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms

United Nations Educational, Scientific and Cultural Organization [UNESCO] (2024) User empowerment through media and information literacy responses to the evolution of generative artificial intelligence (GAI). UNESCO CI/FMD/MIL/2024/3. Available at https://unesdoc.unesco.org/ark:/48223/pf0000388547

UNICEF (2021a) Digital misinformation/disinformation and children. Available at https://www.unicef.org/innocenti/media/856/file/UNICEF-Global-Insight-Digital-Mis-Disinformation-and-Children-2021.pdf

UNICEF (2021b) Digital Literacy in Education Systems Across ASEAN: Key insights and opinions of young people. Available at https://www.unicef.org/eap/media/7766/file/Digital%20Literacy%20in%20Education%20Systems%20Across%20ASEAN%20Cover.pdf

United Nations (2018) UN Resolution A/RES/73/27 Developments in the field of information and telecommunications in the context of international security. Available at chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://documents.un.org/doc/undoc/gen/n18/418/04/pdf/n1841804.pdf

United Nations (2019) United Nations Strategy and Plan of Action on Hate Speech. Available at https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf

United Nations (2022) Countering disinformation for the promotion and protection of human rights and fundamental freedoms (A/77/287) 12 August. Available at https://digitallibrary.un.org/record/3987886?ln=en&v=pdf

United Nations (2023) Our Common Agenda Policy Brief 8 Information Integrity on Digital Platforms. Available at https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf

United Nations (2024) United Nations Global Principles For Information Integrity: Recommendations for Multi-stakeholder Action.Available at https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf

United Nations General Assembly [UNGA] (2021). Disinformation and freedom of opinion and expression. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan. A/HRC/47/25. Available at https://documents.un.org/doc/undoc/gen/g21/085/64/pdf/g2108564.pdf?token=tgTLkICQgazk1ji4Va&fe=true

United Nations General Assembly [UNGA] (2022). Countering disinformation for the promotion and protection of human rights and fundamental freedoms. Report of the Secretary General. A/77/287. 12 August. Available at https://documents.un.org/doc/undoc/gen/n22/459/24/pdf/n2245924.pdf?token=hfIkoXdUIYZOoHeJjJ&fe=true

United Nations Human Rights Committee (2011). General Comment no. 34. CCPR/C/CG/34. Available at https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

United States National Association for Media Literacy Education [NAMLE] (2023a) NAMLE Core Principles on Media Literacy and Education. Available at https://namle.org/resources/core-principles/

United States National Association for Media Literacy Education [NAMLE] (2023b) NAMLE Core Principles on Media Literacy and Education: Implications for Practice. Available at https://namle.org/wp-content/uploads/2023/05/IFPs.pdf

Van der Meer et al. (2023) Can Fighting Misinformation Have a Negative Spillover Effect? How Warnings for the Threat of Misinformation Can Decrease General News Credibility. *Journalism Studies, Vol. 24, N. 6.* Available at https://www.tandfonline.com/doi/pdf/10.1080/1461670X.2023.2187652

Vosoughi S et al. (2018) The spread of true and false news online. *Science Vol 359, Issue 6380*. Available at https://www.science.org/doi/full/10.1126/science.aap9559

Yang A (2024) Mark Zuckerberg apologizes to parents at online child safety hearing. *CBN News.* Available at https://www.nbcnews.com/tech/social-media/mark-zuckerberg-apologizes-parents-online-child-safety-hearing-rcna136578

Yadav K et al. (2021) Countries have more than 100 laws on the books to combat misinformation. How well do they work? *Bulletin of the Atomic Sciences*. Available at https://thebulletin.org/premium/2021-05/countries-have-more-than-100-laws-on-the-books-to-combat-misinformation-how-well-do-they-work/

YLE News. (2022) Supo: Spreading fake news on behalf of foreign power should be illegal. YLE News. Available at https://yle.fi/a/3-12565990

Yun Chee F (2024) Meta to set up team to counter disinformation, AI abuse in EU elections. *Reuters*, 26 February. Available at https://www.reuters.com/technology/meta-set-up-team-counter-disinformation-ai-abuse-eu-elections-2024-02-26/

Ward M et al. (2019) Formative Battles: Cold War Disinformation Campaigns and Mitigation Strategies. Wilson Center. Available at https://www.wilsoncenter.org/sites/default/files/media/documents/publication/cold_war_disinformation_campaign.pdf

Wardle C and Derakhshan H (2017) Information disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe Report DGI(2017)09.* Available at https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

White House (2018) National Cyber Strategy of the United States of America, Trump White House. Available at https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

White House (2020), Executive Order on Preventing Online Censorship, Trump White House. Available at https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-preventing-online-censorship/

White House (2021) Executive Order on the Revocation of Certain Presidential Actions and Technical Amendment, Biden White House. Available at https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/14/executive-order-on-the-revocation-of-certain-presidential-actions-and-technical-amendment/

White House (2023) Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Biden White House. Available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

World Economic Forum [WEF] (2024) Global Risks Report. Available at https://www.weforum.org/publications/global-risks-report-2024/digest/

World Standards Cooperation (2024) Standards collaboration on AI watermarking, multimedia authenticity and deepfake detection. Available at https://www.worldstandardscooperation.org/standards-collaboration-on-ai-watermarking-multimedia-authenticity-and-deepfake-detection/