Vilnius University Faculty of Law

Department of Public Law

Ugur Petcu,

II study year, International and European Law Programme Student

Master's Thesis

Freedom from Disinformation as an Emerging Human Right

Laisvė Nuo Dezinformacijos Kaip Besiformuojanti Žmogaus Teisė

Supervisor: Assist.Dr. Donatas Murauskas

Reviewer: Dr. Johanas Baltrimas

Vilnius

2022

ABSTRACT AND KEY WORDS

Increasing use of social media has brought the emergence of disinformation. Both national and international authorities responded with regulatory solutions to protect the people from its adverse effect. However, coming up for an effective remedy while avoiding interference with freedom of expression provides a great challenge. This study approaches the notion of disinformation from a human rights perspective and focuses on national and international policies to define, monitor and respond. It discloses types of regulatory responses with limitations of freedom of expression and identifies the importance of a universal definition for an effective and fair regulation.

Keywords: disinformation, misinformation, freedom of expression, national security, freedom from disinformation

LIST OF ABBREVIATIONS

BBC: British Broadcast Company

ECJ: European Court of Justice

ECHR: European Convention on Human Rights

ECtHR: European Court of Human Rights

EP: European Parliament

EU: European Union

EU Charter: European Union Charter on Fundamental Rights

EU CoPD: European union Code of Practice on Disinformation

EU DSA: European Union Digital Services Act

GDPR: General Data Protection Regulation

HLEG: High Level Expert Group on Fake News and Online Disinformation

ICCPR: International Covenant on Civil and Political Rights

NetzDG: Network Enforcement Act (Germany)

OSCE: Organization for Security and Co-operation in Europe

UDHR: Universal Declaration of Human Rights

UK: United Kingdom

UN: United Nations

UNGP: UN Guiding Principles on Business and Human Rights

UN HRC: United Nations Human Rights Committee

US: United States

WHO: World Health Organisation

TABLE OF CONTENTS

INTRODUCTION

As of April 2020, approximately 66% of humanity uses the internet (International Telecommunication Union, 2022). This indicates that over 5 billion people have access to the world's largest data network. This network, where more than 3 billion e-mail exchanges are made every day, is often used as the key instrument to reach information quickly and to communicate with other people in the world. Bringing together more than 4 billion users (Statista, 2022), social media platforms can be seen as the largest networks where people communicate en masse and interact with each other. While in the past only editors had the authority to decide which message will be delivered to the public through the media, in today's world, potentially anyone and any content can reach millions of people instantly. As a result, social networks play an important role in promoting participation of people in political decision-making, providing different perspectives to public debate, promoting transparency and oversight of governmental decisions, and bringing together advocates from all over the world on matters concerning human rights, the rule of law and democracy.

However, the increasing use of these platforms has brought with it a new dilemma. While these platforms enhance and strengthen the individual's freedom of expression by providing information and facilitating discussions on political, social and economic issues, they also create an ideal environment for the comprehensive and systematic dissemination of false information, that is, disinformation (HLEG, 2018).

Although the concept of disinformation has been evaluated under different names throughout history (Taylor, 2013), the action itself and the underlying motivations have remained. While this motivation may be for economic benefit, such as creating fabricated news for profit (BBC,2019), It could also aim to create political instability, as in the massive disinformation campaigns prior to the annexation of Crimea (Treisman, 2016). Through the advancement of information technologies, the ever-increasing use of social platforms where anonymous content can be shared, such as social media platforms, provided the optimum environment to individuals and groups who try to spread false information for political, social and economic purposes, therefore involuntarily increasing its effects. In this context, many anonymous accounts are used for disinformation activities, and through these accounts, popular social media platforms such as Facebook and Twitter have become dangerous structures that spread misleading/harmful content. These contents, which consist of false information such as manipulated photos and audio recordings, can spread rapidly among the masses (Bennett and Livingston, 2018).

Moreover, once news based on inaccurate information begins to spread, even if these contents are later revealed to be unfounded, the spread of false information is often not stopped and its influence can continue for a while to affect people's ideas and opinions (Marwick and Lewis, 2018: 44).

Following the various controversial social media activities during the 2016 elections of USA and Brexit referendum (UK Interim Report, 2018), one of the biggest manifestations of the rising disinformation threat was observed in the recent Covid-19 pandemic. News that did not contain any basis, such as that drinking alcohol destroys the virus in the body, that the virus was produced in the laboratory, that the disease was a designed experiment, became widespread on social media. To such an extent that the United Nations expressed this rapidly spreading disinformation campaign with the concept of "infodemic", as a reference to the resemblance in the spread of the pandemic and its effect on people (WHO, 2020).

As a reflection of the rising threat of disinformation, regulations regarding the monitoring and control of information disseminated on the internet have not been delayed. However, finding appropriate responses to disinformation presents a challenge in itself. On this point, the lack of a common consensus on the concept of disinformation provides a wide range of interpretations to these responses, which is difficult as it is vulnerable to misuse and the scale and nature of the issue is debated in the absence of adequate data and academic study. While countries such as Germany, Hungary, and Singapore have enacted laws that impose fines or imprisonment for acts of publishing and disseminating disinformation online (Bradshaw, 2019), some countries have taken preventive methods such as raising awareness of the public against disinformation (Helm and Nasu, 2021). On the other hand, the European Union has taken a self-regulatory approach in cooperation with social media companies on disinformation. (EU CoPD, 2018)

In terms of human rights, the biggest challenge in responding to the spread of false information stems from the restrictive nature of these measures in relation with the freedom of expression. Guaranteed by many international agreements such as the ICCPR, freedom of expression, which includes rights such as the right to access information, is considered as the foundation for "political debate, truth-finding, social cohesion, avoidance of censorship, and self-development" therefore is one of the most indispensable rights in a contemporary civilization (Eskens et al., 2017, p1). However, although this right provides comprehensive protection, it is not unlimited when conditions regarding its limitation and provisions such as the prohibition of hate speech are considered (ICCPR Art. 20).

Aim of the research

This research aims to emphasise the growing necessity for a universal definition and concept by evaluating regulative responses regarding the spread of disinformation through a human rights perspective.

Objectives of the research

For achieving this aim, the following objectives are set:

1.To analyse the attributes and the spread of false information by comparing different terminologies and its historical uses to properly reflect the phenomenon in order to propose an inclusive terminology at the international level.

2. To disclose the provisions that can shape or limit the regulatory responses against disinformation-based activities within the framework of the rights and freedoms given under the international human rights conventions.

3. To evaluate regulative responses of national/international administrations and social media platforms against the spread of disinformation in relation with their effectiveness and applicability.

Description of Research Methods

For achieving the aim and objectives, the terminological analysis and description methods permitted conceptualization of the notion of disinformation given in the thesis. In fact, beginning with the description of the act of disseminating false information, it was possible to identify the special attributes and gain a thorough comprehension of the various elements. These descriptive methods, along with an analysis of the existing terminology, illuminated the intricacies and particulars of the phenomena to be addressed.

The comparative method and the legislative analysis used regarding the regulative responses of the states and social media platforms are particularly relevant to emphasise the lack of a uniform approach. In essence, only by comparing the existing laws of public and private entities it can be shown that there are concerns with the efficiency of such separate approaches. These techniques were also employed in the evaluation of limitations on the act of disseminating false information under freedom of expression.

Structure of the research

This study analyses the concept of disinformation and the regulative responses under three chapters. The first chapter deals with the act of spreading false information with its reasons and consequences, it presents an inclusive terminology for today's disinformation-based activities.

The second chapter examines the provisions of freedom of expression in international human rights conventions, also specifies the limitations of these provisions and presents examples from case law. The third and last chapter analyses the reactions to disinformation under separate headings, while also giving place to international disinformation policies.

Delimitation

This study focuses on the correlation between regulatory responses regarding dissemination of false information and freedom of expression. Considering the scope of information that a single study can process, only the legal texts within the United Nations and the European Union will be examined among the international human rights standards that define the freedom of expression. The same attitude applies to the processing of international disinformation policies. In the third part of the study, the types of regulation are given with their main lines and prominent features. In order to obtain more detailed information, it is recommended to examine the academic studies on each species separately.

Originality

There are numerous academic research on violations of freedom of expression, especially ECtHR and ECJ jurisprudence, however most of this research are on illegal speeches such as hate speech and defamation. Therefore, when evaluated in terms of the concept of disinformation and misinformation. limitations of freedom of expression is a relatively new field. In addition, it should be taken into account that many regulative responses against the dissemination of false information covered in the thesis have been made in the last few years.

Important sources:

This research focuses on the regulative responses against the spread of false information. Thus, the primary sources consist of the international human rights conventions, such as ECHR and ICCPR. UN General Comment No.34 and ECtHR Guide on Article 10 were also used as the legitimate interpretations of the provisions establishing freedom of expression and other related rights. In the examination of EU and UN disinformation

policies, which are evaluated within the scope of this thesis, the reports and legislations prepared by the relevant institutions, such as the HLEG report also constitute important sources.

CHAPTER I – UNDERSTANDING THE CONCEPT OF DISINFORMATION

1.1. The Act of Spreading False Information

Establishing a clear definition of disinformation that satisfies the normative question of what should be considered illegal is instrumental for the development of appropriate measures. However, there is not much consensus in the academic literature on how to identify and operationalize many disinformation-related phenomena. The widely diverse applicability range of the laws that apply to disinformation reflect this definitional vagueness. Although The terms "Misinformation" and "Disinformation" are often used in a mixed manner with varying and conflicting definitions, these terms are conceptually different (Guess and Lyon, 2020, p.10). The act of spreading false information constitutes the defining element as the common denominator of all disinformation and disinformation-based activities. It would therefore be consistent to treat the extent and effects of the dissemination of false information throughout the action itself. In addition, it is possible to distinguish other terms shaped around this descriptive element by their characteristic features.

1.1.1. Disinformation

The concept of disinformation, which is defined as deliberately misleading information, comes from the Russian word "*dezinformacija*" in 1949 (Karlov and Fisher, 2013). Typical examples of disinformation include deceptive advertising, government propaganda, photos with bad content, fake documents and maps, internet fraud, and fake websites. From this point of view, the concept of disinformation, which lies on its basis, is extremely dangerous. Because the fact that the news contains false and misleading information in the fields of economy, medicine, politics and culture can cause serious emotional, financial and even physical destruction in people (Fallis, 2014).

According to the statements of Soviet intelligence officer Sergei Tretyakov, in the years when the Cold War was effective, disinformation activity was a method frequently used by the Russian Intelligence Agency (KGB). In particular, the news about nuclear war (Agbedo and Krisagbedo, 2014).

Luciano Floridi, one of the first thinkers to have an idea about the concept of disinformation, provided a statement in 1996 as "disinformation occurs when the information processing phase is faulty". Seven years later, Floridi's analysis that "when the semantic content is wrong, it gives rise to misinformation, if the source of the misinformation is made consciously, there is disinformation here" can be interpreted as repairing the previous deficiencies regarding the concept of disinformation. In 2009, scientist Don Fallis defined disinformation as "misleading information". Don Fallis does not see the lies that are hard to believe in misleading information as disinformation, and comments that "in disinformation, information may be wrong, but disinformation can be made with correct information". As of 2010, philosopher Brian Skyros defines "systematic use of misleading information" as disinformation (Fallis, 2014).

Perhaps the most modern definition of disinformation today is the definition made in the HLEG' final report. According to HLEG final report (2018, p.5), disinformation can be inclusively defined as "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit".

With the spread of new communication technologies, people have the opportunity to obtain information faster. In the field of media, individuals have gained the opportunity to learn about the developing events on the internet every second. Therefore, the information age, in which computers and the internet are tools, has also revealed new disinformation methods. For example, giving information to deceive internet users by imitating the websites of reputable institutions and news sources, creating fake or modified images with image processing techniques (software such as Photoshop) have made disinformation widespread and facilitated (Fallis, 2009).

The following strategies are generally used to disseminate information intended for disinformation:

(1) Causing negative comments by presenting misleading information to the targeted audience,

(2) Drawing the image of supporting and defending the real information to the targeted audience and making comments that will invalidate that information and distance it from the target,

(3) Adding additional distorting information to the targeted segment (Agbedo and Krisagbedo, 2014).

The types of disinformation used to mislead the masses by distorting the truth can be listed as Disinformation which is usually effective in the public or military sphere (Fallis, 2009):

(1) Disinformation is usually effective in the public or military sphere,

(2) Disinformation is deception, usually using carefully planned and technically advanced methods. For example, deliberately changing the content of news sites such as Yahoo News and The New York Times, which were hacked as a result of hackers' attacks, with false information is also a type of disinformation. Disinformation can be applied by interfering with the content of the website without hacking the websites.

(3) Disinformation often comes across in written and verbal form, but another common method is manipulated photos.

(4) Disinformation spreads more rapidly through television, newspapers or the Internet. Disinformation, which generally means "false information", has been on the rise in the field of new media lately. In this context, disinformation can target specific individuals or organisations as well as general-purpose use with mass media.

Furthermore, some incorrect information may exist due to human error or, on the contrary, in the purpose to entertain the audience, such as satire or parody. In this case, the intention of parody must be either to entertain or to encourage thinking and public debate within public opinion. Although the mentioned parody might consist of provocative content, it cannot be considered as a threat. The ECtHR accepts parody as a “form of artistic expression” (Ziembiński v. Poland 1799/07, 2016).

1.2. Propaganda as an Early Format

When we look at the first examples of propaganda in a serious sense, although it has its roots in the 16th century, this concept actually sprouted together with the formation of societies. It is possible to find traces of propaganda since the formation of the first societies. So much so that, over time, the creation of tales or stories full of grudge and hatred for tribes and societies that societies saw as enemies for themselves was put forward to intimidate and discourage their enemies, and in these epic stories, each society showed itself as a hero, while the cruel ones became their perceived enemies. (Qualter, 1980, p. 257). Harold Lasswell, who has academic studies in the field of political communication, makes the following definition of this concept in his work The Theory of Political Propaganda, which he wrote in 1927; “it is the management of collective attitudes through the manipulation of important symbols (p. 627).” After Lasswell's explanation of this

concept, he also opens the word attitude in the same paper and states that there is a tendency to behave according to certain value models. Symbols, on the other hand, stated that there are gestures, mimics, colours that replace words and meanings, and references that can correspond to different meanings in different cultures (Lasswell, 1927, p. 627).

With the twentieth century, World War II and the clash of nations' ideologies drove them to engage in propaganda on differing shapes. However, during the conclusion of the cold war, most scholars switched away from propaganda, assuming that it had become an old phenomenon as Florian Zollman (2019) called "The Marginalisation of Propaganda".

Looking at the existing definitions, it should be noted that similarly, the following common emphases are highlighted:

(1) Propaganda is limited to one-way communication.

(2) It is imposing, it is desirable to instil the desired idea and does not offer different ways.

(3) When necessary, every way is considered permissible for the intended purpose.

(4) It is created in a planned and deliberate way.

(5) Propaganda contains information, and even if it is wrong or incomplete, it is information after all, and propaganda is based on persuasion, and persuasion occurs through repetition.

Considering these points, it can be said that propaganda includes actions and purposes that are defined as disinformation today. However, the following detail should be noted here, while disinformation is based on the spread of misinformation, it does not deal with the type of information as long as it is for propaganda purposes. The Joint Declaration on Freedom of Expression and Fake news, Disinformation and Propaganda (2017), defines notions of Disinformation and Propaganda in a similar manner. Point 2.6 of the declaration provides the only difference as: “State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false, or which demonstrate reckless disregard for verifiable information.” Although the distinction in the given definition is rather vague, it means that the perpetrator is either in the knowledge or should have known (within the reason) that the information he disseminated is not true and still continues to spread it or is not concerned with the accuracy of the information he disseminates. In all these given cases, the person essentially reveals that he does not care about the accuracy of the information. In addition, ICCPR considers any kind of propaganda promoting war as “War Propaganda” and strictly prohibits it (ICCPR (1966, Art. 20 para 1).

1.3. Factor of Intent

When we evaluate the types of dissemination of false information in terms of "intent", we will encounter two basic forms in this field: disinformation and misinformation.

1.3.1. Misinformation

Misinformation is defined as inaccurate and misleading information conveyed by persons or institutions in an unintentional manner (HLEG, 2018). According to the Cambridge dictionary, the prefix "mis-" indicates that an action has been performed improperly or poorly. Typically, misinformation refers to false information that is not noticed as such and is disseminated without the goal of spreading misleading information and deceiving people. However, the consequences might still be damaging. When information is shared in its original form, it has a lasting effect on one's memory. So much so that even if the information is later found to be wrong, the initial information remains in the memory of the person. Therefore, with this attitude, the person trusts and supports the information containing misinformation (Ecker et al., 2014).

Taking a different approach to misinformation, Christopher Fox (1983) comments the relationship between knowledge and misinformation as “knowledge does not have to be true; therefore, there is no harm in the fact that the information is correct, and there is no harm in the fact that it is wrong; Therefore, misinformation is a type of knowledge.”. From this point of view, misinformation is information even if it is false, and it can also be informative. Misinformation, which is a type of information, can be briefly expressed as incomplete information. In addition, hiding, distorting, and falsifying information can be included in the concept of misinformation. Because information that seems incomplete and irrelevant can still be accurate, up-to-date and informative (Agbedo and Krisagbedo, 2014).

Different factors influence the expansion of the phenomenon of misinformation, which contemporary societies cannot prevent in parallel with the Internet's siege of the whole world. Some researchers who have a say in this field (Lewandowsky et al., 2012) suggest the following views on which strategies are applied and what is effective in the dissemination of false information is built on rumours and fictions. Although societies have struggled with rumours and their repercussions over the years, unfortunately rumours are one of the important sources of misinformation. For example, the opinion that the Bush administration did not take any measures to stop the September 11 attacks, according to most US Democrats in 2006, is the most striking of the strong rumours created (9/11

Commission Report, 2004). Although the media, by definition, aims to inform the public, it also helps to spread misinformation systematically.

1.4 Magnitude and Impact

Disinformation is not a new concept. However, social media platforms' excessive growth has influenced the public lifestyle including news consumption and distribution patterns (Kalsnes 2018, p. 1). As people lock themselves into a narrow chamber of information (echo chamber), where they access information only through the prism of their own ideology or political position they tend to believe the content they see without questioning (Sunstein, 2017).

Considering the atmosphere of uncertainty and social and economic difficulties it brings, the Covid-19 pandemic process has provided a very fertile ground for disinformation activities (Bayer et al, 2021, p. 100). In addition, social measures to combat the virus, such as quarantine and social distancing, are putting many people in social and financial distress (Bayer et al., 2021, p. 101). Studies show that such financial hardships and lack of freedom lead people to believe in disinformation about COVID-19 (Oleksy et al, 2021, p. 1).

When the effects of disinformation are examined, it is seen that it causes the public to approach the news containing accurate information with suspicion in the long run by decreasing the trust in the media. Prolonged exposure to inaccurate and manipulated and distorted information can damage deep-rooted democratic institutions by increasing scepticism and even apathy in individuals (Barbera, 2020). The social impact of this problem was seen during the 2016 Brexit referendum, USA presidential elections in the same year and the recent COVID-19 pandemic, where various false claims widely spread (Tucker et al., 2018, p. 1). The impact of online disinformation is much greater in developing countries with underlying instability, poor media landscape and inexperienced internet users. Because in such countries, online disinformation can easily turn into real-life violence.

Given the possible advantage of a foreign state in creating instability or showcasing its superiority, governments are one of the main sources of disinformation. This situation may be illustrated by Russia's organised disinformation campaigns in Ukraine prior to the annexation of Crimea (Treisman, 2016). However, due to the complexity of these actions and a lack of data, determining the actual impact is challenging (European Parliament Report, 2019).

States can also benefit from dividing their own society for gaining more support or maintaining power. Various studies associate the act of disinformation with political polarisation (Tucker et al, 2018, p.4) (Guess and Lyons, 2020, p.10). This type of government disinformation is often followed by censorship. Serious concerns have been raised earlier within the EU that Hungary is running such campaigns (Bayer, 2021, p 46). Some scholars argue that the rise of populist leaders such as Donald Trump around the world has contributed to this problem (Barbera, 2020, p.34). This is because these individuals' used disinformation as a convenient tool in their polarization campaigns, which they implemented in order to wear down the current political regime during election periods and to strengthen their voter base during their own rule (Hameleers, 2020, p.146.).

Financial incentives can also be cited as another reason for the rise of online disinformation. As the investigations revealed the Macedonian citizens employed to fabricate and share false information who are earning three times more of the average wage, disinformation can also be seen as a profitable business (BBC, 2019).

CHAPTER II – FREEDOM OF EXPRESSION

2.1. Human Rights Conventions

2.1.1. Universal Declaration of Human Rights

The Universal Declaration of Human Rights (UDHR), adopted by the United Nations General Assembly in 1948, is the first international human rights document that recognizes freedom of expression as one of the fundamental human rigts, emphasizes that it should be protected, and announces it in more than 500 languages. (UDHR,1948, preamble) In the aftermath of the Second World War, there was a consensus of the United Nations not to repeat this indescribable atrocity around the world. With this declaration, which is a natural consequence of this war, world leaders have decided to strengthen the UN Charter by consecrating and promoting human rights guarantees everywhere (Brown, 2016, p. 1).

The General Assembly stated in 1946 that freedom of expression is a fundamental human right and this freedom is the cornerstone of all other freedoms that the United Nations has a duty to protect. In accordance with this statement, two years later, freedom of expression was shapened in Article 19 of the UDHR. It declares Freedom of expression as follows: "Everyone has the right to freedom of opinion and expression; This right includes the freedom to hold opinions without interference and to seek, receive and disseminate information and ideas by any means and regardless of borders." (UDHR, 1948, Art. 19).

Although UDHR does not legally bind signatory countries, it has been accepted as a statement of principles that will serve the countries of the world as a "common standard of achievement" (Hannum, 1998, p. 145). Since its proclamation in 1948, UDHR has served as the most important benchmark by which the content and standard of compliance with human rights is measured worldwide, and has become increasingly moral, political, and legal. Accordingly, the view that the rights and freedoms it enshrines, including the provision on freedom of expression, gain legal force as general principles of customary international law (Howie, 2018, p.12). According to Eide and Alfredsson (1999), the performance and even legitimacy of governments should be measured according to UDHR standards.

2.1.2. International Covenant on Civil and Political Rights

Although the foundations of freedom of expression and the related right to obtain information take shape with article 19 of the UDHR, it will be the International Covenant on Civil and Political Rights (ICCPR, 1966) that shapes the limits of this right in detail. Article 19 of the ICCPR expands on the basic principles outlined in Article 19 of the UDHR and combines them with the duties, responsibilities, and restrictions that shape the exercise of freedom of expression.

The first paragraph of Article 19 of the ICCPR guarantees one's right to have an opinion. As for the second paragraph, ICCPR shapes the freedom of expression to include “the freedom to seek, receive and impart information and ideas of all kinds…regardless of frontiers”. In addition, the aforementioned right can be exercised privately in the form of an oral, written, printed or artistic product or by other means of communication of its choice.

From this, we can see the foresight and efforts of article 19 of the ICCPR for the exercise of freedom of expression in all kinds of platforms that may be used in the future, as well as the possibilities at the time it was issued. As a result, the framework of international human rights is still relevant today and applies equally to modern communication technologies such as the internet and social media. Furthermore, article 19 of the ICCPR allows Governments to:

(a) Safeguard the freedom of expression by abstaining from interference with its exercise, (b) exercise due diligence to protect this right or to prevent, punish, investigate and redress violations of freedom of expression,

(c) taking positive or proactive measures to enforce or allow the realisation of this right. (UN HRC General Command No.34, 2011)

The ICCPR has not established a judicial body with the power to interpret its provisions in a legally binding manner, however, it has a committee on how its various articles should be interpreted, such as Article 19, which defines freedom of expression. The Human Rights Committee is a treaty body founded under Article 28 of the ICCPR to monitor the implementation of the Covenant. The Committee consists of eighteen members who are nationals of the signatory parties as well as "persons of high moral character and recognized competence in the field of human rights" (ICCPR, 1948, Art. 28). While the Committee's interpretation of Article 19 is not legally binding, it can provide crucial guidance. Furthermore, the statements of United Nations special rapporteurs are particularly useful in understanding the right to free expression.

While freedom of expression is crucial for democracy and the exercise of other freedoms, as discussed in the previous section, it is not absolute. The third paragraph of article 19 of the ICCPR recognizes that the exercise of the right to freedom of expression may be limited in certain exceptional circumstances. However, as will be discussed in the next section, the restriction must be prescribed by law, serve legitimate purposes, and deemed necessary in a democratic society.

2.1.3. European Convention for the Protection of Human Rights and Fundamental Freedoms

The first document that made human rights shaped by UDHR legally binding was the European Convention on Human Rights, signed in 1950. It has been created both by the interpretations of UDHR’s texts by the European Court of Human Rights and by the work of the Council of Europe (ECHR, 1950, preamble). Besides the additional protocols, the council has implemented different resolutions and recommendations that expand the scope of the Convention. Article 10, paragraph 1 of the Convention defines freedom of expression as follows: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television, or cinema enterprises."

Together with the case law of the European Court of Human Rights (ECtHR), the establishment of above article undoubtedly contributed to a greater degree of respect for freedom of expression and free media in the member states of the Council of Europe. As a

reflection of this, ECtHR's Handyside v. The United Kingdom decision reinforces this idea by noting that “freedom of expression constitutes one of the main foundations of a democratic society and constitutes one of the basic conditions for the progress of this society and the self-realization of every individual” (Handyside v. UK, EctHR, No. 5493/72, 1976.). In its decision, the Court stated that freedom of expression, defined in Article 10 of the Convention, covers "not only information or ideas that are favourably received or viewed as inoffensive, but also those that are offensive, shocking or disturbing"(para. 49) Indeed, despite their troubling nature, such statements are imperatives of diversity, tolerance, and open-mindedness, without which no democratic society would exist. (Sunday Times v. the United Kingdom, EctHR, No. 6538/74, 1979)

Article 10 on protection of freedom of speech not just applies via written or spoken words, but also through visuals (Muller and Others v. Switzerland, EctHR, No.10737/84, 1988. para 35), images (Chorherr v. Austria, EctHR, No. 13308/87, 1993), activities (Steel and Others v. United Kingdom, EctHR, No. 851/1058, 1998), and even cultural assets (Khurshid Mustafa and Tarzibachi v. Sweden, EctHR, No. 23883/06, 2008). In some cases, even a dress may fall within the scope of Article 10 (Stevens v. United Kingdom, No. 11674/85). Moreover, as the Court noted, Article 10 safeguards not just the content of ideas but also the manner in which they are expressed. Therefore, all kinds of compositions such as paintings, films, novels and satyrs are protected in accordance with this article (Bychawska-Siniarska, 2017, para 33).

2.2. Limitations of Freedom of Expression

Like most fundamental rights, freedom of expression is not unconditional. Its execution may be limited to protect conflicting public interests and the rights of others as it will be discussed in the following. However, there are some differences between human rights documents such as regarding the meaning of legitimate aim, which is one of the reasons for restricting freedom of expression. For example, the UDHR only permits limits on freedom of expression “for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society” (UDHR, 1948, Art. 27). On the other hand, ICCPR and ECHR regulate in more detail the reasons of limitation.

Despite the differences, all human rights instruments stipulate that state measures to limit freedom of expression must be justified by (1) being prescribed by law, (2) having a legitimate aim, and (3) being necessary for a democratic society (ICCPR, 1966, Art. 19(3)).

In other words, there is an obligation to abide by the principle of proportionality. This three-stage test has been developed in the jurisprudence of international judicial bodies, particularly the UN Human Rights Committee and ECtHR. In most cases, the burden of proving that all three requirements are met rests with the state (Bychawska et al., 2017, p.33). When international judicial bodies find that the state has failed to prove one of the three requirements, they will not examine the case any further and will decide that the relevant interference was unjustified and thus violated the freedom of expression.

It should be noted from the outset that burden of justifying the restrictions imposed on the right to freedom of expression by the state rests with the state, not the individual. In UN General comment 34, the Committee notes that it is the state's responsibility to explain the exact nature of the danger, as well as the necessity and proportionality of the specific action implemented. (UN General Comment No. 34, para 35).

2.2.1. Conditions of Limitation

2.2.1.1. Being prescribed by law

According to this requirement, any interference with this right must be based on a clearly formulated law. In principle, this means a written law, sufficiently accessible and predictable, shaped with sufficient precision to allow an individual to regulate his behaviour. The Human Rights Committee, in its General Comment 34(para 26), states that the law limiting freedom of expression should itself “be compatible with the aims and objectives of the Convention”. For example, laws that violate the right to non-discrimination or sanction corporal punishment do not meet the criterion of legality because they violate the provisions of the Convention. Accordingly, in order to remain permissible, the limitation must not only be "prescribed by law" but also provided by law that is substantially in accordance with the requirements of the Convention.

In Kivenmaa v. Finland (Comm. No. 412/1990), the applicant, who is an author organizes a protest that featured a banner criticizing the human rights record of a president who is a guest at that time. The protest took place outside the presidential palace, and the author was prosecuted for holding a "public meeting" without prior permission. The author claimed that the law of public meetings did not apply to him, as he was expressing an opinion rather than holding a public meeting. The state referred to the grounds on which freedom of expression could be limited under Article 19 (3) of ICCPR and justified the limitation in general. The Committee, however, rejected the State's claim, underlining that the applicant's freedom of expression was not legally limited.

Similarly, the phrase "prescribed by law" should also be interpreted broadly. According to ECtHR, this does not only indicate that the act under dispute must have a legal basis. The accessibility and substance of the law under discussion should not be considered independent of this requirement (ECtHR Guide on Art. 10, 2022, para 68). It should also be noted that in terms of accessibility, the court considers that the announcement of the law in the official gazette is sufficient (Güzel v. Turkey, EctHR, No. 29483/09, 2016.). In the case of Akcam v. Turkey (EctHR, No. 27520/07, 2012.), it is emphasized that the criminal law provisions should clearly state the definition and scope of the crime in question. This is for the sake of preventing expansion of the state's margin of appreciation in the prosecution of relevant offenses which can potentially be abused through arbitrary enforcement.

Finally, it is important to note that as law tends to lag behind rapidly changing technological and social conditions, it is not uncommon for measures limiting freedom of expression to be based on a legal framework that is not internet-adapted and does not offer the required level of predictability.

2.2.1.2. Making the Intervention for a Legitimate Purpose

Once the regulatory requirement is met, the next test is whether the restriction is for one of the permissible instances of the legitimate purposes identified. Accordingly, any restrictions on freedom of expression can only be imposed in order to achieve one of the legitimate aims broadly stated in the relevant human rights instruments. In other words, these justifications can only be the protection of the reputation and rights of others, national security, public order, public health or morals. Considering all the circumstances in which freedom of expression may be limited, there is no other possible legitimate purpose other than the ones mentioned above (ICCPR, Art.19(3a)(3b))

However, the given term of "*ordre public"* is a broader concept than 'public order' and includes general welfare and even public policy. ECHR Article 10(2) provides examples by setting seven possible restrictions permitted under of these *ordre public* criteria (excluding the reputation and rights of others corresponding to ICCPR Article 19(3)(a). The legitimate grounds for restricting freedom of expression in Article 10 (2) of ECHR are stated as follows: "in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the respect of rights or reputations of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary."

2.2.1.2.1. Rights or Reputation of Others

The first ground of limitation specified in human rights documents is the "rights or reputations of others". The scope of the term "others" may be interpreted as "other persons individually or as members of a community"(UN General Command No. 34, para 28). Moreover, this ground of limitation is closely related to Article 17 of the ICCPR (1966), which sets “No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to the unlawful infringement of his honour and reputation." In essence, limiting freedom of expression to "respect the reputation of others" can be interpreted as right to remain free from unjustified interference on others' reputations (Gunatilleke, 2020, p 134). As an opposing view, Manfred Nowak's work can be cited. Nowak (2005, p 462) argues that even humiliation, slander and insults that are not considered in protective scope of Article 17 may be restricted by law under Article 19 (3) “to the extent that it may be necessary to respect the reputation of others”. While other scholars such as Nowak have argued that such limitations should be evaluated under the principle of proportionality, an interpretation that the term "reputation" in Article 19 (3) extends beyond the right protected in Article 17 is controversial, both conceptually and normatively.

The preliminary case law of the ECtHR shows that the phrase “rights and reputation of others” as a justification within the legitimate reasons of limiting freedom of expression is not interpreted in the light of "the Right to Respect for Private and Family Life" (ECHR Art. 8, 1950). The case of Lingens v. Austria (EctHR, No. 9815/82, 1986), is one of the important cases that indicates how the court interpreted the conflict between freedom of expression and rights and reputations of others. The applicant is an Austrian journalist and editor of a magazine. In an article he wrote, he criticises the Socialist Party leader for trying to form a government with the Liberal Party. According to him, the head of the Liberal Party is a politician who served in the SS brigade during the Second World War. Therefore, seeking an alliance with the political party led by such a person is "opportunistic" and "immoral impudence". Upon these statements, the President of the Socialist Party sued the journalist Lingens on the grounds that the crime of defamation had been committed, and the Austrian court fined Lingens 15,000 shillings.

In this case, which was brought to ECtHR by Lingens, it is claimed that the relevant event constitutes a violation of Article 10. As a result of its examination, the court determined that the intervention met the conditions in pursuit of a legitimate aim in terms of protecting the rights and freedoms of others and prescribed by law. However, the main debate here is

whether interference was necessary in a democratic society. As a result of the examination, the court concluded that the freedom of expression stated in the first paragraph of Article 10 of the ECHR is one of the essential features of a democratic society and, taking into account the importance of Linens' statements, it concluded that there was a violation. As a reason for this decision, he stated that freedom of the press is one of the important tools that enable the public to have an opinion about the thoughts and behaviours of political leaders. For this reason, it was said that the limit of reasonable criticism should be wider for politicians than for other people, because unlike other people, politicians have opened their words and actions to the close scrutiny of both the public and journalists, so they should approach the comments and criticisms made about them more tolerantly. According to the Court, the important point here is to make a decision by comparing the requirements to protect the reputation of politicians in limiting their expressions with the benefits of public discussion of political issues.

Following Lingens case, ECtHR began to interpret the term reputation as part of the right to respect for private life. The development of the Court in this direction is clearly seen in the Feldek v. Slovakia (ECtHR, No. 29032/95, 2001), Lesnik v. Slovakia (ECtHR, No. 35640/97, 2003) and Radio France v. France cases (ECtHR, No. 53984/00, 2004.). In Chauvy v. France (ECtHR64915/01 2004, para 70), the Court states that: “The right to protect one's reputation. . . It is one of the rights guaranteed as an element of the right to respect for private life."

The Court's acknowledgment of reputation as a factor of Article 8 has a significant influence on the freedom of expression. This affects the court's approach to defamation cases. Prior to the Lingens trial, the court used proportionality analysis to rule on freedom of expression and reputation cases. When the Court applies proportionality analysis, the margin of appreciation given to member states is narrowed and the level of scrutiny by the Court becomes much more stringent. The Court tends to use balancing rather than proportionality, after the notion of individual dignity has been accepted as part of Article 8. In such a case, the principle of subsidiarity would apply, this implies that states will have a wider margin of appreciation, and the ECtHR's oversight will be reduced.

ECtHR's interpretation of satire and parody can be seen in Ziembiński v. Poland (ECtHR, 2016, 1799/07). In the relevant case, Maciej Ziembiski, a local newspaper owner and journalist, wrote a series of articles and columns in 2004 criticising a local official's intention to expand quail farming in the Radomsko Region. The article refers to the local official who wrote the plan as "nonsense", "posing", and "foolish officer", but did not

explicitly give a name. The article also refers to his "bosses" (himself) as "boring". Upon this article, the employees of the relevant facility filed a lawsuit accusing Ziembiński of defamation through the mass media, and they allege that the statements in the article discredit them in public. The District Court found Ziembiński guilty of defamation committed through the media.

Although not named in the article, the court concludes that government officials could be easily identified, given previous articles in the paper. While acknowledging the journalist's right to openly criticise public authorities, the court determines that Mr. Ziembiski's article constituted a clear breach of free expression and professional ethics and should be penalised. Following the decision, Ziembiski filed an appeal with the European Court of Human Rights, alleging that his defamation conviction infringed his freedom of expression under Article 10 of the ECHR. In its decision, the Court found that there had been a violation of Article 10 of the ECHR, but specifically criticised the domestic court's failure to adequately consider the satirical nature and ironic implications of the article. The Court finds “the use of sarcasm and irony is perfectly compatible with a journalist's exercise of freedom of expression” (para 44) and states that “while any individual who takes part in a public debate of general concern … must not overstep certain limits, particularly with regard to respect for the reputation and rights of others, a degree of exaggeration or even provocation is permitted; in other words, a degree of immoderation is allowed.”

2.2.1.2.2. National Security

One of the most prominent reasons given by nations for interfering with freedom of expression is national security. However, since no precise definition is provided for the concept of "national security", the state becomes the ultimate determinant of what constitutes national security. In most cases, the concept is broadly defined and open to manipulation by states. Yet there are cases where exposure of media abuse in the national security sector has led to reforms and ultimately to greater security. Case regarding the Pentagon documents in the USA can be counted among the examples (New York Times Co. v USA, SCOTUS, 403 US 713, 1971).

The ICCPR's Siracusa Principles on Limitation and Derogation of Provisions (1985) corresponds to another source that states the criteria to be taken by states in order to eliminate risks and dangers related to national security. According to the Siracusa Principles, the existence of emergencies that endanger the life of the whole or part of the country constitutes the first condition necessary for the suspension of freedom of

expression. In this context, the suspension of rights and freedoms due to national security is subject to various conditions in all cases where risks occurring in entirety, or a part of the country endanger public health, public peace, and the effective and healthy functioning of institutions. Among these, there are criteria such as the states' official declaration of a state of emergency, the proportionality of the limitation and the legal regulation of its scope, and the fact that the measures are open to the inspection of independent judicial institutions. In addition, it should be ensured on a legal basis that the measures will automatically be lifted once the situation that created the state of emergency disappears.

These principles define a legitimate national security justification as “only when they are taken to protect the existence of the nation or its territorial integrity or political independence against force or threat of force. “. (UN Siracusa Principles…, 1985, para. 29) The following articles state that restraint of national security "cannot be invoked as a reason for imposing limitations to prevent merely local or relatively isolated threats to law and order." Similarly, the UN Special Rapporteur on Freedom of Expression (CN.4/1995/32, 1995, para 48) has repeatedly stated that "“For the purpose of protecting national security, the right to freedom of expression and information can be restricted only in the most serious cases of a direct political or military threat to the entire nation”.

The UN human rights committee interprets the concept of national security sometimes flexible. Mukong v. Cameroon is a case that should be considered as an example (Comm. No. 458/1991). Mukong, a Cameroonian journalist, was arrested for making an interview critical of the government and allegedly subjected to inhumane treatment during detention and held in poor conditions. In addition, he claimed that he was arrested for speaking against the two parties, saying that multi-party democracy is more correct for Cameroon. He was tried and found guilty by the local court for "misleading the national and international public opinion". Following this decision, Mukong made an application to the UN Human Rights Committee and stated that his freedom of expression had been violated. The Cameroonian government claimed that the prosecution of Mukong as a result of his activities and statements in this case complied with the conditions of limitation set out in article 19, paragraph 3. In this case, the Committee stated that "the prevailing political context and situation in a country at any time must be taken into account" in assessing the exercise of the right to freedom of expression. Following this statement, the committee emphasised the need to "strengthen national unity" in Cameroon. In this case, the Committee admitted that it "indirectly justified its actions on the basis of national security and/or public order" on the grounds of "the political context of the country and the ongoing

struggle for unity". In this case, the committee did not specify a country's political context and whether the ongoing struggle for unity had a direct relationship with "national security" or "public order". However, he defined the justification of “preserving and strengthening national unity under difficult political conditions” as a “legitimate” aim to restrict freedom of expression. Therefore, the Committee has accepted a purpose that is not fully listed among the limitations set out in paragraph 3. As a result, no breach of Article 19 of the ICCPR occurred. It should also be emphasised that the Committee identified "preserving and strengthening national unity under difficult political conditions" as a "legitimate" reason for limiting freedom of expression. Therefore, the Committee accepted a purpose not fully listed in Article 19(3).

Considering the fact that national security is a vital interest for all states, the European Court of Human Rights gives member states a wide margin of appreciation where national security is presented as a justification for restrictions on freedom of expression. While the Court and the Commission have great discretion in assessing states' threats to national security and deciding how to combat them, they may be out of context and insufficiently equipped to detect real threats to the state concerned. The Court is, therefore, inclined to require national bodies to verify that any threat is of reasonable cause (Spielmann 2012, p.402).

In this respect, glorification of terrorism is seen by many countries as a threat to national security and becomes a justification for restricting the right to freedom of expression. The term incitement, glorification, or praising terrorism is different from what could be defined as promoting terrorism. In this matter, incitement can be defined as the direct encouragement of one's criminal acts with the aim of inspiring another person whom the speaker does not know to perform the act. In support of this, The ECtHR has also accepted, within its case-law, that expressions that incite "hatred, revenge, blame or armed resistance" or "violence, armed resistance or insurrection" fall outside the scope of Article 10 (Murray, 2009, p.340).

In the case of Zana v. Turkey (ECtHR, 69/1996/688/880), the question of national security was raised. Zana, a former mayor of a Turkish city, was imprisoned for making the following comment in a newspaper interview: “I support the PKK national liberation movement; On the other hand, I am not in favour of massacres. Anyone can make mistakes; the PKK accidentally kills women and children…”. Convicted of "endangering public safety", Zana was sentenced to 12 months in prison (para. 26) . Due to the PKK being a terrorist organisation, the government invoked national security and public security as a

legitimate aim provided for in paragraph 2 of Article 10. In this case, the Court considered the situation and circumstances surrounding the applicant's statement as a whole. Accordingly, the Court noted that the applicant's statement was made at a time when tensions in southeast Turkey were escalating, and the PKK was carrying out attacks that claimed many lives. Considering the career of Zana, who had served for a time as mayor in southeastern Turkey, the court therefore considered that her statements were of greater importance. Therefore, considering the margin of appreciation offered to the states on the grounds of national security in addition to these reasons, the court ruled that the said decision of the defendant Turkish state was proportionate to the legitimate aims pursued and that there was no violation of freedom of expression.

However, in Incal v. Turkey (EctHR, App. No. 41/1997/825/1031), although the applicant, like Zana, had been convicted under the same article of the Criminal Code, the ECtHR found that his freedom of expression had been violated. Because in Incal v. Turkey, the court was of the opinion that "unlike Zana, it is unlikely that conditions will worsen an already escalated situation". The applicant's statement was also considered a political criticism of the government, which enjoyed stronger protection than other forms of expression. Cases such as Urper and Others v. Turkey (EctHR, App. No. 14526/07, 2010) and Director Duman v. Turkey (EctHR, App. No. 15450/03, 2016.) can also be cited as examples where the issue of national security is put forward as a justification for restricting the right to freedom of expression.

To sum up, when states resort to the protection of national security or public order as a legitimate justification allowing restriction of freedom of expression, the Court will consider subjective factors, i.e., the characteristics of the applicant himself, as well as objective factors such as the context and circumstances surrounding the interference. This reveals that among the details the Court took into account during its assessment was the applicant's impact on society. Because the applicant's profession, whether he/she has the title of public official or duties such as the chairman of a non-governmental organization/foundation will affect the strength of the statement made. On the other hand, the media organ where the speech took place and the extent of interaction caused by the news will also be considered.

2.2.1.2.3. Public Moral

The protection of public morals is also stated as a legitimate aim justifying the restriction of freedom of expression. Unlike other justifications for limiting freedom of expression, there is no definite method to determine if a certain activity effectively threatens public morality. The inability to determine the scope and boundaries of public morality can lead to arbitrary restrictions on fundamental rights (Perrone, 2014, p361). However, the main question that can be raised is who sets the public moral standards as the justification for limiting the right to freedom of expression.

It is also important to note that, unlike other reasons for restriction, the UN Human Rights Committee and the European Court of Human Rights have a different understanding of public morality. The traditional view of the ICCPR is that blasphemy law may provide a legitimate purpose regarding Article 19 (3), which allows for limitations on freedom of expression in the name of public morality (Cox, 2020, p36). However, in its General Comment 34, the committee concluded that “prohibition of non-respect for a religion or other belief system, including libel laws, is incompatible with the ICCPR” (UN General Comment 34, para 48). According to this interpretation, laws prohibiting blasphemous speech can never be justified ECHR on the grounds of protecting public morals and are fundamentally incompatible in relation of human rights.

On the other hand, the ECtHR does not consider general moral justification to be illegitimate in nature. Rather, it applied the well-known margin of appreciation doctrine of discretion to the concept of "common morality" and recognized the signatory states as those most qualified to concretely set the context of the content. In Handyside v. the United Kingdom, the Court stressed that, in the case of expressing an opinion on the exact content of moral requirements and the 'necessity' of the 'prohibition' or 'punishment' they devised to meet them, State authorities, by virtue of their direct and sustained contact with the fundamental powers of their country, usually in a better position than a judge (Handyside v. United Kingdom, EctHR, No. 5493/72, 1976, para 52).

This situation has been confirmed in subsequent cases, and it has also been clearly stated that the margin of appreciation accorded to the contracting states regarding general morality is quite wide (Muller and Others v. Switzerland, EctHR, No.10737/84, 1988. para 35: Open Door and Dublin Well Woman v. Ireland, EctHR, No. 14235/88, 1992, para 68).

The general moral code emerges, along with other interests, as a justification for limiting individual rights in cases involving religious discourse, such as in Tatlav v. Turkey(50692/99, 2006 para 21), these rules are seen in cases dealing with hate speech, where state officials consider the content of some public speeches to incite “hatred and hostility on religious and racial grounds”, as in the Gunduz v. Turkey (35071/97, 2012, para 28) and Erbakan v. Turkey (59405/00, 2006, para 46) judgments.

2.2.1.2. The Necessity of Interference in a Democratic Society

The third part of the three-stage test is the criterion that restriction is necessary for a democratic society. According to this requirement, interference with freedom of expression is acceptable only if there is a proportional relationship between the interference and its legitimate aims. In General Comment no 34(para 22), the UN human rights committee states that " restrictions must be applied only for those purposes for which they were prescribed and must be directly related to the specific need on which they are predicated."

In the Sunday Times v. the UK (ECtHR, No.13166/87, 1991, para 50), The ECtHR emphasised that for an intervention to be regarded necessary in a democratic society, it must correspond to a "pressing social need" that is relevant, appropriate, and proportionate to the legitimate goal pursued. The necessity of a “pressing social need” mentioned in this formula seems to be related to the weight and importance of the goals pursued. In other words, it is not enough that the interests served by the restriction of freedom of expression are legitimate, there must also be "oppression".

Limitations should also be proportional. This means that the advantages resulting from the interference must outweigh the disadvantages in terms of limiting the exercise of freedom of expression. The prerequisite restricts government officials from exercising their powers, obliging a balance to be struck between the means used and the result intended. After assessing the existence of social need pressure, the ECtHR usually assesses the nature and severity of the interference. It has also repeatedly stressed that member states should use the least restrictive measure.

CHAPTER III – RESPONSES TO DISINFORMATION

3.1 European Union Disinformation Policies

3.1.1 EU Code of Practice on Disinformation

Aiming to hold the 2019 European parliamentary elections in healthier conditions, the European Commission established the High Level Experts Group (HLEG) at the beginning of 2018. The aim of the group was to make recommendations on policies that could be developed against the increasing disinformation on the internet. In its report, HLEG defined the concept of disinformation as “false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit”(EU CoPD, 2018, p. 5). Following this report and European Commission Communication "Tackling online disinformation: a European approach" (April 2018), the Commission cooperated with various institutions and organisations to take regulatory measures against disinformation. Among them were corporations such as Google, Facebook and Twitter, as well as institutions such as the Association of European Communication Agencies.

The EU Code of Practice on Disinformation (October 2018), which was drawn up as a product of all these efforts, took the definition of disinformation made in the HLEG report and, with some additions, brought it to the following form:

"verifiably false or misleading information, which, cumulatively,

(a) "Is created, presented and disseminated for economic gain or to intentionally deceive the public”

(b) "May cause public harm", intended as "threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security” (EU CoPD, 2018, preamble para 2).

Another important detail regarding the definition was that misleading advertisements, satire and parodies were not included in the concept of "Disinformation" (EU CoPD, preamble). According to the CoPD, exposing individuals to disinformation through misleading or inaccurate information is a big concern, because the survival of modern democracies rely on public discussions that allow its citizens to freely interact with each other and share their ideas. The signatories state that they recognize and approve all these statements while also acknowledging their recognition of how important and delicate the balance between freedom of expression and limiting the spread of otherwise lawful content. While the Act states that trade associations that sign this code are not liable on behalf of their members,

it does express the necessity to keep their members fully informed of these Rules and to encourage them to abide by it or to respect its principles as appropriate.

Disinformation campaigns often represented a business or were aimed at generating profits. Therefore, the initial first commitment to "reduce the revenues of disinformation suppliers" was directed at regulating ad placements and ad monetization incentives. (EU CoPD, 2018, Commitment II.A, para 1)

The second commitment deals with political and issue-based advertising. Since there is a transparency principle in political advertising, platforms must clearly indicate the candidate or party that makes the payment, together with the amount of the advertisement, apart from the content. At this point, it should be noted that among the signatories, there are institutions such as Facebook that comply with this commitment, as well as those that do not.

The third commitment, titled integrity of services, essentially stipulates a regulation for fake accounts and states that the distinction between human and artificial intelligence accounts should be ensured. However, it contains a restrictive regulation: "Signatories should not be prohibited from enabling anonymous or pseudonymous use of accounts" (EU CoPD, 2018, Commitment II.C para 6).

The fourth commitment, entitled "Empowering consumers", advises that a policy of balance should be struck between the goals of freedom of expression and prevention of disinformation. In general, it emphasizes the need for signatories to invest to improve cooperation, and also emphasizes that these technologies should give content priority to "relevant, authentic, and authoritative" information. However, it is unclear how these technologies will be used jointly and a standardization will be achieved.

The fifth commitment includes arrangements to strengthen the research community. The critical issue in this area is data privacy and access to data. At this point, it is emphasized that the analysis of data is of critical importance in the understanding and treatment of disinformation, and the importance of collaboration between platforms and the research community is emphasized. The regulation also states that various signatories operate differently due to their technologies and target audiences, thus enabling different approaches for signatories to fulfil the provisions contained in this regulation. As a result, potentially creates a range of flexibility for signatories regarding the appliance of the act. This is also one of the justifications of why the law adopts a self-regulatory approach and allows signatories to choose which provision to commit to. However, the possibility of choosing different commitments for each signatory further restricts the effectiveness of the

implementation of the agreement, which may have difficulties in terms of legal binding due to its already limited self-regulatory approach. At this point, due to the self-regulatory approach, the stipulated measure will not go beyond "inviting the signatory to withdraw from the Law" (EU CoPD title V. Signatories) if one of the signatories of the agreement does not comply with the provisions. For this reason, the implementation of the provisions of the agreement will be left only to the will of the signatories, which will lead to the dysfunction of the agreement. This situation must have drawn the attention of the Sounding Board, as in the opinion they gave in September 2018, it is stated that the disinformation law does not contain "a common approach, clear and meaningful commitments, measurable objectives", and also does not have a monitoring or enforcement tool, and therefore cannot be classified as self-regulation (The Sounding Board's Unanimous…, September 2018).

The commission, which frequently brought up the possibility that self-regulation would yield an insufficient result before the 2019 elections, and highlighted that a regulatory approach could be adopted, expressed the need for additional regulations, although it did not maintain the same insistence after the elections. The first annual report, which highlights the need to develop the platform in terms of consumer empowerment and collaboration with the research community, was published in October 2019. Accordingly, the data and search tools which are provided to the research community were "episodic and arbitrary." (EU CoPD, Annual Reports, 2019, p 11). However, the general assessment stated the code "has provided an opportunity for greater transparency into the platform's policies on disinformation as well as a framework for structured dialogue to monitor, improve and effectively implement those policies" (Commission Staff Working Document…, 2020).

Insufficient organization among stakeholders, particularly between verification bodies and the research community, in the 2019 report led to the establishment of the Global Alliance for Responsible Media (GARM) to increase collaboration. GARM, which started out with 16 members three years ago, continues to work today within the framework of a community with 122 members and its founding goals (World Federation of Advertisers, 2022).

3.1.2 EU Digital Services Act: End of Self-Regulation

Together with its sister legislation the Digital Markets Law; EU Digital Services Act (DSA), which forms a single set of rules valid throughout the EU, aims to end the self-regulation period in which technology companies set their own policies against disinformation and publish their own transparency reports (EU Digital Services Act,

2022/2065). The regulation, which holds major internet platforms such as YouTube and Facebook responsible for societal harm caused by the use of its services, imposes heavy fines for violations, forcing them to be more active in combating the spread of illegal content across the EU. It also introduces two new classifications of very large online platforms (VLOP) and very large online search engines (VLOSE) and imposes stricter obligations on companies in these categories due to their high risk of spreading illegal and harmful content, including disinformation.

The DSA, which contains more than 300 pages of regulations, consists of detailed rules from the legal obligations of technology companies to the responsibilities of the EU and member states. Some of the key regulations can be listed as follows:

(1) Regarding illegal content: The regulation, which imposes a ban on monitoring general content across the EU, forbids platforms from systematic inspections in a way that undermines freedom of expression. However, it stipulates that in violation of national and EU law, the digital service provider must take swift action and remove the content (DSA, 2022/2065, Art. 8).

(2) Users right to challenge content moderation decisions: Users are required to be provided with detailed explanations in case of account blocking and content removal/demotion by platforms. It also gives users affected by these decisions the right to appeal and offers the option to seek out-of-court remedies if necessary (DSA, 2022/2065, para 58).

(3) Increased transparency regarding online advertising: Platforms are required to clearly articulate how content moderation and algorithmic recommendation systems work in their terms of service, and to offer the option of an alternative content recommendation system that is not based on profiling. (DSA, Article 15)

(4) Limitation on deceptive designs and targeted advertising: Limiting the ad designs that can be accepted to deceive and manipulate users, DSA prohibits the profiling of individuals using "sensitive" features such as sexual orientation, religious belief, and the use of ads targeting children in particular.

(5) Reporting: Regulation requires platforms to report annually on content moderation. These reports, which require that all automatic systems used for content control be explained with error rates and working principles, should also include the rate of

complaints, how the complaints were responded to, and the number of parties who made the complaint about the removal of the content.

(6) On the largest platforms and system risks: Recognizing that the biggest risk to fundamental rights, elections and public health comes from the biggest platforms, legislators foresee heavier obligations for platforms with more than 45 million users in the EU (DSA Article 33 para 1). Accordingly, platforms with the specified features are obliged to formally evaluate what measurable steps can be taken to prevent the above-mentioned risks in society for all their products, including their algorithmic systems.

(7) Mandatory data access (Only for external review): Platforms are obliged to share all their internal data with EU and member states officials, especially with independent auditors, in order to audit their work on self-assessment and risk reduction. The fact that these data are also accessible to academic researchers is a positive detail for future studies.

(8) New authority and responsibilities for the European Commission and state authorities: Coordination will occur between new national and EU-level authorities. The Commission will have direct monitoring and enforcement powers over the main platforms and search engines, as well as the power to impose fines of up to 6% of their global revenue (DSA Article 52, para 3). The Commission may potentially impose regulatory costs on platforms to assist them fund their own enforcement efforts.

DSA, published in the EU Official Journal on 27 October 2022, is in effect as of 16 November. Each member state is required to nominate its own digital services coordinator by 17 February 2024, when the law becomes fully applicable throughout the EU. In this role, the coordinator is an independent regulator responsible for liaising with the commission and enforcing the law on smaller platforms established in his country. This period starts earlier for the largest platforms and search engines, although the law will apply across the EU from February 2024. It allocates four months to the platform after the European Commission's designation to comply with the DSA. The recently established European Center for Algorithmic Transparency will begin its review of VLOP and VLOSE compliance by summer 2023. Lastly, considering that DSA is planned as a single legislation valid throughout the EU, the need for revision of many national regulations, such as the German NetzDG, seems inevitable.

3.2 Regulative Responses for disinformation

The fact that disinformation causes information conflict in society has recently increased the awareness of disinformation. This has led to the development of different regulatory provisions to control the content of information accessible to people.

The target of these different regulatory approaches ranges from internet users to the platforms where information is produced and disseminated from sources of disinformation (Helm and Nasu, 2021, p.302). The following section discusses information correction, penal sanctions and content moderation as examples of regulatory responses to the spread of disinformation.

3.2.1 Information Correction

Information correction formations are a mechanism to minimise the negative impact of online disinformation, in which certain content is publicly explained and proven to be false through a digital platform (Bontcheva, 2020, p. 66). From a freedom of expression perspective, this is known to be the least intrusive form of regulation because it is based on the assumption that individuals are rational enough to seek the truth when faced with questionable information (Helm and Nasu, 2021, p 314).

The idea of creating an online portal to correct disinformation is quite common in many countries such as China, Croatia, Italy, and Pakistan. According to Helm and Nasu (2021, p. 315), the government of Democratic Republic of Congo has employed young people to track and report the spread of false information circulating in social media. Their decision can presumably be linked to recent wave of disinformation about the Ebola epidemic. Their teams are even operating on WhatsApp to respond such disinformation activities with accurate information.

Singapore has adopted a more invasive approach to data correction. Under the Online Frauds and Manipulation Protection Act 2019(10/2019), government authorities may issue a “correction order” to demand an individual or internet tool service provider who has made a misrepresentation to submit a correction notice in a specific form. However, many social science researchers show that information correction alone is not an effective mechanism to minimise the harmful effects of online disinformation (Man-pui, 2017, p 35).

3.2.2 Penal Sanction

Penal sanction is known as an effective way to minimise the negative impact of disinformation in the digital environment (Helm and Nasu, 2021, p. 322). It aims to prevent the creation and sharing of fake news with a proactive approach. In this way, manipulated information cannot be spread to the public. Recently, countries such as Malaysia, Germany, France, Hungary, Singapore, Russia, and Nigeria have enacted laws that provide fines or imprisonment for acts of posting and disseminating disinformation online (Bradshaw et al, 2018, p 8: Mchangama, 2019, p 1).

The main problem of such a regulatory solution is its incompatibility with the normative requirements of the right to freedom of expression. The use of criminal sanctions to prevent the creation and dissemination of false or misleading information may be considered disproportionate due to the potential for arbitrary practice. The UN and various organisations such as Organization for Security and Co-operation in Europe (OSCE), emphasise the importance of this situation in their joint declaration titled Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda (2017) as follows: "General prohibitions on the dissemination of information based on vague and ambiguous ideas, including "false news" or "non-objective information", are incompatible with international standards for restrictions on freedom of expression, as set out in paragraph 1(a), and should be abolished." They also pointed out that their criminal laws were unduly restrictive and should be repealed. However, some scholars argue that a penal code prohibiting the fabrication or spread of disinformation can be drafted with specific and sufficient precision to address legitimate interests. (Helm and Nasu, 2021, p. 325)

Essentially, a penal regulation against the spread of disinformation promises to be effective when the positive aspects such as the deterrent factor it will bring and the ability to prevent disinformation before it spreads are taken into account. The person who prepares the disinformation content is aware of the fact that the information he spreads is false and acts in a self-interested manner, which creates criminal liability. Rather than rejecting such measures on the basis of uncertainty; The solution may be to include several safeguards, such as a proportionality test, a threshold of harm, and an intent to defraud requirement to limit the extent to which criminal prosecution is used to protect legitimate interests (Helm and Nasu, 2021, p 325). However, the greatest solution that can prevent arbitrariness in this area would be to unite in a universal definition of disinformation or to reach a binding consensus that reveals the principles of disinformation. The joint declaration of the United

Nations against disinformation is an important step that guides the international recognition of this threat and how it should be approached, but it lacks a legally binding nature.

3.2.3. Content Moderation

In order for a form of regulation to be introduced to combat disinformation to be effective, it must have mechanisms including the removal or blocking of the disinformation content in question. This requirement, on the other hand, creates the need for two separate legal regulations: The legal definition of the "illegal" content to be removed, and the determination of the responsibility of the platform on which this content is spread in terms of removing and blocking the content. (Sander, 2021, p. 163)

3.2.3.1. Content Restriction Laws

International human rights law imposes an exceptional obligation on states to prohibit certain expressions and to limit their dissemination, to prevent the harm that various forms of expression may cause. For example, Article 20(2) of the ICCPR (1966) states that "Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law." According to this regulation, restricting the content of disinformation, which includes elements corresponding to discrimination, hostility or incitement to violence, is incompatible with the prevention of freedom of expression and is compatible with international law. However, that restriction is not exempt from the burden of meeting the requirements of Article 19/3 of the ICCPR.

3.2.3.2. Liabilities of Intermediaries

Internet intermediaries such as Google, Facebook, Apple, Microsoft, Twitter, Linked In and Yahoo! are playing an increasingly important role as platforms that effectively shape norms and boundaries for how users can form or express ideas and encounter information (Spano, 2017, p. 162). As mentioned earlier, such websites allow an unlimited number of people to post and share their ideas and opinions, and let others learn and comment on their opinions. However, the same platforms are also used to spread disinformation, propaganda, hate speech and other forms of illegal content that ultimately endanger citizens' human rights. Therefore, countries around the world have established different types of agent liability regimes that impose liability on social media platforms when illegal content is distributed by their users (Sander, 2021, p. 160).

Illegal content of users, including disinformation, is evaluated under three different liability models which are strict liability model, safe harbour model and broad immunity model (MacKinnon, 2015, p. 39). However, before making a descriptive analysis of these three models, it is important to answer the question of whether the international human rights regime (often focusing on state action) can be applied to such non-state actors (social media companies). It is also vital to determine whether social media should qualify as a technology company or a media company. The answers to the two questions above determine to what extent social media companies are responsible for any illegal content produced and disseminated by their users.

3.2.3.3. Social Media Companies and their Responsibility to Protect Human Rights

Considering the human rights law, historical process and general principles, it can be said that it is established to be applied to countries. (Taylor, 2016, p.3). However, as transnational corporate actors gained enormous power and wealth, their negative impact on human rights began to inflame debate at the United Nations.

This process, which gave birth to initiatives that offer advice for corporations to operate in accordance with human rights, resulted in the adoption of the UN Guiding Principles on Business and Human Rights (UNGP) by the UN Human Rights Council in 2011.

Companies must "respect" human rights, which indicates they must avoid violating human rights and respond to any negative human rights violations in which they are engaged, according to these guiding principles. Although they do not have all of their human rights obligations, institutional actors are expected to refrain from adversely affecting the enjoyment of human rights and to offer remedies if rights are violated (Aswad, 2019, p. 39.). In general, the UNGP expects corporations to establish human rights policies, actively interact with external stakeholders in analysing human rights concerns, apply due diligence to analyse possible threats to human rights, and implement measures to avoid human rights violations.

Furthermore, the Human Rights Council emphasizes that “a company's obligation to respect human rights is a worldwide norm that exists independently of states capability and/or willingness to fulfil their human rights obligations and does not diminish those obligations.” (Organisation for Economic Co-operation…, 2011, para 37). The UN special rapporteur on freedom of expression noted in his 2018 thematic report that while these Guiding Principles are non-binding, The dominance of social media corporations in public

life throughout the world gives a compelling justification for their acceptance and deployment. (Kaye, 2018, para. 11).

3.2.3.4. Digital Constitutionalism and Accountability Gap

The development of information technologies has not only challenged the fundamental rights of individuals such as freedom of expression, protection of personal data and privacy of private life. The new technological infrastructure, built under the guidance of liberal ideas, specifically empowered technology companies that control social media platforms to perform some semi-public functions (De Gregorio, 2021, p. 41).

Therefore, this discussion is no longer limited to the field of private law, but shifts to the field of public law, especially constitutional law. As it is known, one of the main aims of modern constitutionalism is to protect fundamental rights by limiting state powers. From a constitutional law perspective, the concept of authority has traditionally been executed by public authorities (Wehba, 2019, p. 27). However, as information technologies have advanced, a new kind of special authority has arisen, a digital authority. Because of the lack of accountability, most constitutions are ineffective when it comes to the abuse of such special authority (De Gregoio, 2021, p. 42).

To fill this legal gap in the digital age, the concept of "digital constitutionalism" has been developed to reinterpret how the use of authority should be limited (legitimized). According to Redeker, digital constitutionalism is "a set of political rights, governance norms, and the exercise of power over the Internet." (Redeker, 2018, p 303) As such initiatives are adopted, countries around the world are limiting the powers of online platforms through the implementation of legal tools aimed at increasing the degree of transparency and accountability in online content management.

3.2.3.5. Technology and Media Companies

Social media platforms often see themselves as technology companies rather than media companies. When Mark Zuckerberg, CEO of Facebook, was asked by the US Congress about the nature of his company, he replied, "When people ask us if we're a media company … my understanding of what the heart of what they're really getting at is, 'Do we feel responsible for the content on our platform?' The answer to that, I think, is clearly yes." (Kelly M., National Public Radio, 2018). He responded to this question in an interview in 2016 by saying "No, we are a tech company, not a media company" (Segreti G., Reuters). However, academics such as Andras Koltay and Barrie Sander do not agree with such a

definition. “While social media companies are not responsible for producing the bulk of the content that appears on their platforms, they do make important decisions regarding both the permissibility and visibility of such content” they claim (Koltay, 2019, p. 157: Sander, 2020, p. 939). Therefore, Facebook, Twitter and other platforms are media companies that make billions of editorial decisions every day by selecting, sorting and removing pieces of content. In other words, Koltay and Sander equate content control with traditional editorial functions.

Such a distinction would have important consequences from a legal/regulatory perspective. Because if the platforms are considered media or broadcasters, they will have more responsibility such as a newspaper, television or radio broadcaster for illegal content uploaded and shared on their networks. However, if we consider them as a technology company, they are not responsible for any content published on their website, just as the postman is not responsible for what is written in a letter or what is written in a librarian's book.

3.2.4 Models of Intermediary liabilities

Many governments in these regions, including Europe, North America, parts of Southeastern Asia, and Latin America, have laws specifically covering intermediary obligations (Mac Kinnon, 2015, p 40). In other regions, particularly in Africa, governments are considering legal provisions regarding intermediary liability (Zingales, 2020, p 1). In general, where such regimes exist, there are three patterns of intermediary liability: strict liability, "safe harbour" (conditional liability), and broad immunity. The particular standards and intricacies of these models differ by jurisdiction and are specified by governments and clarified further by the courts (Hong, 2015, p 369)

3.2.4.1. Strict Liability model

In countries that follow this model, the agent is responsible for the content of third parties. The only option of the intermediary, who wishes to avoid any sanctions within the framework of this responsibility, is to proactively monitor, filter and remove such illegal content (Hong, 2015 p 369). In case the intermediaries fail to fulfil their liabilities, they face a variety of sanctions, including revocation of business licences and criminal penalties (Callamard, 2020, p 205). An example of this type is the Tort Act, which came into force in People’s Republic of China in 2010. According to article 36/3 of this law, online intermediaries are held jointly responsible with the internet user if they are in the knowledge that the internet user using the services they offer is using their systems to commit harmful

acts, but they do not take action to stop this illegal activity (Hong, 2015, p 371). Delfi v. Estonia (2015, 64669/09) provides an example from Europe. In this case, the European Court of Human Rights holds that an online news website was liable for allegedly defamatory comments posted by an anonymous user.

3.2.4.2 Safe Harbour Model

According to this concept, intermediaries have the option to be excluded from responsibility for third-party material with the condition of providing they remove illegal content after obtaining information. Unlike the strict liability model, the safe harbour approach does not obligate agents to actively monitor and control information in order to elude legal accountability. The application tool of this concept is known as the "notice and takedown" mechanism (Hong, 2015, p.372). Notice and takedown procedure generally refer to the process of removing content by the agent upon notification of a person whose rights have been violated for a specified period of time.

This method can be found in both regional and national policy laws. One of the widely known notice-and-takedown method is found in the United States Digital Millennium Copyright Act (DMCA, 105-304, 1998). The procedure in the DMCA only deals with copyright-infringing content (Buiten, 2020, p.140). An important element of the DMCA is that it allows internet intermediaries to block access to illegal material as long as they act in good faith.

Article 14 of the European Union E-Commerce Directive (2000/31/EC) describes a notice and takedown mechanism but does not include a direct definition. Under this provision, hosting providers may benefit from liability exemption if, after learning about the illegal nature of the information, they take action to quickly remove the information or block access. The provision applies to any illegal or infringing content, including disinformation.

In practice, however, notice and takedown mechanisms are often introduced in industry-specific regulations and often only cover uniform content. This is an indication that most of the existing Notice and Takedown mechanisms only apply to copyright infringement. Over time, some countries have expanded the scope of the mechanisms provided to include other types of violations. Hungary, for example, has expanded the mechanism to cover the personality rights of minors in addition to copyright infringement (Electronic Commerce and on Information Society…2001, Art. 13(1). Germany’s recently passed NetzDG Act also requires social media providers to remove explicitly illegal content (e.g., hate speech) upon complaint (Sander, 2021, p. 174.).

The notice and takedown procedure is subject to criticism mainly for three reasons:

(1) It can be abused; that is, if there is no legal process, such as an opportunity to challenge the takedown request, this mechanism may allow agents to remove the content immediately after receiving the notice, rather than investigating whether the content is actually illegal (Sander, 2021, p 178). For example, if Facebook is notified that certain content considered to be disinformation has been posted on its platform, Facebook is allowed to remove that content without examining whether it is genuine disinformation. Legal content may be censored as a result.

(2) It can promote collateral censorship, i.e. governments may try to indirectly regulate content creators based on various measures to influence the content moderation practices of social media companies, rather than directly regulate speakers and broadcasters. For example, according to the Twitter transparency report 20, Turkey sent more than 47,000 legal demands for content removal related to nearly 200,000 accounts and this number is only from the second half of 2021(Turkey, Twitter Removal Requests, Jul-Dec 2021).

(3) This mechanism is the improper transfer of legal authority to the private sector (Article 19, 2013 p 16). In other words, the content shared on their platforms with the intermediaries provides a quasi-judicial position in assessing its legality. However, intermediaries are not qualified to replace the courts and do not have the democratic legitimacy to decide the fate of an individual's freedom of expression.

3.2.4.3. Broad Immunity Model

This model does not hold its affiliated intermediaries responsible for the third-party content they host or disseminate (MacKinnon, 2015, p. 42).

The United States is an example of granting broad immunity to major social media companies and all other intermediaries. Article 230 of the Communications Decency Act (CDA) defines this policy as: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." (47 USC CDA 230(c)(1)).

Unlike other intermediary liability models implemented in Europe, the USA does not force social media companies to comply with any notification and takedown procedures as a reflection of this policy and does not impose any liability. According to John Samples, a member of the Facebook Overcite Board, the main purpose of the American government's broad disclaimer under this policy is to encourage technological development, while

respecting the freedom of expression of social media platforms and users (Samples, 2019, p 4). The point that should not be forgotten here is that the legal regulation (CDA) acts with the expectation that social media companies will create their own content policies in the future (47 USC CDA 230).

While private social media companies are broadly immune to liability for illegal content posted by users on their platforms, the First Amendment of the USA Constitution does not bind these media companies because it only protects citizens from being banned from speaking by the government. Thus, despite this general assumption in favour of freedom of speech, regulatory actions by private social media companies through in-service terms and community guidelines will not constitute positive obligations under the First Amendment (Klonick, 2017, p. 1609).

CONCLUSIONS

1. Disinformation is not a new issue for humanity, throughout history, many individuals or groups have used false information as a tool for political and economic interests. Although this action is evaluated under different classifications today as it was in the past, when it is examined considering its causes and effects, a definition under the name of disinformation reveals an inclusive concept. The only concept that differs in this definition is the concept of misinformation, in which the person disseminating the information is not aware that this information is false. This difference does not mitigate the consequences of the action; however, it reveals that the perpetrator is also the victim. In this process, social media creates a convenient area for disinformation to reach the masses where information can be shared without control and the sharer can remain anonymous. As for the reason why people do not question the accuracy of the information they see on these platforms, lies with various factors such as one-sided consumption of content caused by political polarization and the lack of explanations from reliable sources in times of turmoil such as the pandemic.

2. When evaluated in terms of freedom of expression, the scope of legal protection does not change depending on whether the information shared is true or false. Although this situation largely prevents the restriction of freedom of expression against disinformation, international human rights conventions have left some room for the signatory states. However, the situations that can be restricted within this scope of action present legitimate restriction conditions that are interpreted quite broadly in practice. When all these conditions are considered, it is seen that restrictions on freedom of expression against large-scale disinformation activities can only be made for reasons such as national security and political or military threats that affect the entire nation. Accordingly, in addition to serious crises such as possible civil war and terrorism threat, it can be used to get through the general election periods in a healthier way, in which a controversial political atmosphere prevails. Content that violates the right to respect for private life may be restricted too, but this requires an individual assessment of each restriction case, therefore it is not possible to consider this as a genuine way of limitation.

3. When the regulatory responses to disinformation are considered, it is not possible to talk about an effective and inclusive regulation. Penal sanction can be considered the most effective method in terms of the deterrent factor it creates in the fight against disinformation. However, considering that the concept of disinformation is open to interpretation, it brings with it arbitrary and disproportionate practices by countries and subsequently violations of freedom of expression. This highlights once again the

importance of reaching consensus on a common definition of disinformation in terms of preventing arbitrary practice. Content moderation, on the other hand, places responsibility on platforms where disinformation spreads, giving them an active role in the response to disinformation. However, it is highly controversial how effective, impartial, and human rights-compliant audits can be carried out in the hands of profit-oriented private companies. This situation has also drawn the attention of the EU that the self-regulatory nature of Code of Practice on Disinformation has left its place to the auditors appointed by the member states in the newly accepted DSA.

4. Developing regulatory solutions to combat disinformation is difficult. The regulation must effectively prevent misinformation and its detrimental impacts while upholding the right to free speech. The analysis in this study has shown that any regulation cannot be effective in eliminating the negative effects of false news without imposing some restrictions on freedom of expression. Therefore, it is necessary to strike a balance between combatting disinformation and protecting freedom of expression. The most fundamental element of this balance will be to unite in a universal definition for disinformation under the umbrella of the UN. However, given the current controversial status of this definition, the preparation of a guideline document on limitations would preclude free interpretation of disinformation without defining it directly. Limitations on the concept of national security introduced by the Siracusa Principles in 1984 can be given as an example. Such a move unites not only states, but also social media platforms, which are constantly interacting with disinformation and working with states, on a common ground.

LIST OF REFERENCES

1. Legal normative acts

1.1. International legal acts

United Nations General Assembly. “International Covenant on Civil and Political Rights.” Treaty Series 999 (December): 171, 1966.

United Nations, Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights. Doc. E / CN.4 / 1985/4, 1985.

United Nations Human Rights Council (UNHRC), Guiding principles on business and human rights: Implementing the United Nations “protect, respect and remedy” framework, 2011.

1.2. European Union Legal Acts

European Convention for the Protection of Human Rights and Fundamental Freedoms, Council of Europe, ETS 5, 1950.

European Parliament, E-Commerce Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), 2000.

European Parliament, 19 October 2022, Regulation (EU) 2022/2065 of the on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) https://eurlex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A32022R2065&qid=1666857835014

1.3 International Guidelines and Reports

Council of Europe: Committee of Ministers, Guidelines on Human Rights and the Fight Against Terrorism, 2002. No. 804/2002.

Council of Europe: European Court of Human Rights, Guide on Article 10 of the European Convention on Human Rights - Freedom of Expression, 31 August 2020

European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Tackling online disinformation: a European Approach, Brussels, 2018.

European Commission, Annual self-assessment reports. https://digital-strategy.ec.europa.eu/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

European Commission, Staff Working Document, Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement, 2020

OECD, Transparency, communication and trust: the role of public communication in responding to the wave of disinformation about the new coronavirus, OECD Publishing, 2020.

Bontcheva et al. (2020). Balancing act: countering digital disinformation while respecting fFreedom of expression, International Telecommunication Union (ITU) and United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, 2020, p. 66.

Bychawska-Siniarska, D. (2017). Protecting the right to freedom of expression under the european convention on human rights: a handbook for legal practitioners, Council of Europe Publishing, p.33

Kaye, D. (2018). Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35). United Nations Human Rights Council, Para. 11

Hussain, A. (1995). Report of the special rapporteur on the nature and scope of the right to freedom of opinion and expression, and restrictions and limitations to the right to freedom of expression, UN Commission on Human Rights, CN.4/1995/32

1.4. National Legal Acts

German Netzdg Act, Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken, Federal Law Gazette I, p. 3352 ff, 2017

Hungary Act CVIII (2001). Art 13(1) (HU)

People's Republic of China, Tort Liability Law, 2010, 中华人民共和国侵权责任法 (Zhōnghuá rénmín gònghéguó qīnquán zérèn fǎ). http://english.mofcom.gov.cn/article/policyrelease/Businessregulations/201312/20131200432451.shtml

Singapore, Online Frauds and Manipulation Protection Act, 2019(10/2019) https://sso.agc.gov.sg/Act/POFMA2019

United States of America, Digital Millennium Copyright Act (DMCA), Public Law code105-304, 1998.

United States of America, Communication Decency Act, 47 U.S.C. § 230(c) (1996)

2. Special Literature

2.1. Books, Articles and Reports

Agbedo, CU. (2012). Placards as a language of civil protest in nigeria: a systemic functional analysis of the fuel subsidy crisis. IOSR Journal of Humanities and Social Science, 6, 17-26.

Alfredsson, G. and Eide, A. (1999). The universal declaration of human rights: a common standard of achievement, Martinus Nijhoff Publishing

Aswad, E. (2019). The future of freedom of expression online, Duke L. & Tech. Rev, 3, p. 26-70.

Barberá, P. (2020) social media, echo chambers, and political polarization, Social Media and Democracy: The State of the Field, Prospects for Reform. Cambridge: Cambridge University Press (SSRC Anxieties of Democracy), p. 34–55.

Bayer, J. et al., (2021). Disinformation and propaganda – impact on the functioning of the rule of law in the eu and its member states, European Parliament

Benkler, Y., Faris, R., and Roberts, H. (2018). Network propaganda: manipulation, disinformation, and radicalization in american politics, Oxford Academic https://doi.org/10.1093/oso/9780190923624.001.0001

Bennett, WL. and Livingston, S. (2018). The disinformation order: disruptive communication and the decline of democratic institutions, European Journal of Communication, 33(2), p..122-139.

Bloch-Wehba, H. (2019). Global platform governance: private power in the shadow of the state, SMU. L. REV. Vol 72, p. 27

Bradshaw, S., Neudert, LM. and Nothaft, H., (2018). Government responses to malicious use of social media. Riga: NATO Strategic Communications Centre of Excellence.

Brown, G. (2016). The universal declaration of human rights in the 21st century: a Living document in a changing world, Open Book Publishers, Cambridge

Buiten, MC., de Streel, A. and Peitz M., (2020). Rethinking liability rules for online hosting platforms. Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim, Discussion Paper Series – CRC TR 224, Volume 28, Issue 2, p. 140. http://dx.doi.org/10.2139/ssrn.3350693

Callamard, A. (2019). The human rights obligations of non-state actors. Human Rights in the Age of Platforms, p. 205

Chan, MS. et al. (2017). Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychological Science, 28(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Cox, N. (2020). Justifying blasphemy laws: freedom of expression, public morals, and international human rights law. Journal of Law and Religion, 35(1), 33–60. http://doi.org/10.1017/jlr.2020.11

Daragh, M. (2009). Freedom of expression, counter terrorism and the internet in the light of the UK terrorist act 2006 and the jurisprudence of the european court of human rights, Netherlands Quarterly of Human Rights, Volume 27, Issue 3, p.340 DOI:10.1177/016934410902700302.

Ecker, UK et al. (2014). The effects of subtle misinformation in news headlines. J Exp Psychol Appl. 2014 Dec;20(4):323-35. doi: 10.1037/xap0000028. Epub 2014 Oct 27. PMID: 25347407.

Eskens, S., Helberger, N. & Moeller J. (2017). Challenged by news personalisation: five perspectives on the right to receive information, Journal of Media Law, 9:2, 259-284, DOI: 10.1080/17577632.2017.1387353

Fallis, D. (2009). A conceptual analysis of disinformation. iConference, Chapel Hill, North Carolina, 1-8.

Fallis, D. (2014). A functional analysis of disinformation. Paper presented at the iConference 2014 Proceedings.

Fox, C.J. (1983). Information and misinformation. an investigation of the notions of information, misinformation, informing, and misinforming, Greenwood Press, Westport, CT and London

Global Citizenship Commission. The Universal Declaration of Human Rights in the 21st Century: A Living Document in a Changing World (G. Brown, Ed.; 1st ed., Vol. 2), 2016. Open Book Publishers. http://www.jstor.org/stable/j.ctt1bpmb7v

Gregorio, G. (2021). The rise of digital constitutionalism in the european union, International Journal of Constitutional Law, Vol. 19, Issue 1, p. 41

Guess, AM. and Lyons, BA. (2020). Misinformation, disinformation, and online propaganda, Social Media and Democracy: The State of the Field, Prospects for Reform, Cambridge University Press, Cambridge, p. 10.

Gunatilleke, G. (2020). Justifying limitations on the freedom to manifest religion or belief and the freedom of expression under the ICCPR. University of Oxford. P. 3-294

Hameleers, M. (2020). Populist disinformation: exploring intersections between online populism and disinformation in the US and the netherlands. Politics and Governance,8(1),146-157. https://doi.org/10.17645/pag.v8i1.2478

Hannum, H. (1998). The UDHR in national and international law. Health and human rights, 3(2), p. 145-165.

Helm, RK. and Nasu, H. (2021). Regulatory responses to 'fake news' and freedom of expression: normative and empirical evaluation, Human Rights Law Review, Volume 21, Issue 2, p. 302–328, https://doi.org/10.1093/hrlr/ngaa060

Hong, X. (2015). Regulation of e-commerce intermediaries: an international perspective. Trade Development through Harmonization of Commercial Law, Victoria University of Wellington, Hors Serie, 19.

Howie E. (2018). Protecting the human right to freedom of expression in international law. International journal of speech-language pathology, 20(1), 12–15. https://doi.org/10.1080/17549507.2018.1392612

Huff, M. (2018). Joint declaration on freedom of expression and "fake news," disinformation, and propaganda, Secrecy and Society. 1. 10.31979/2377-6188.2018.010207.

Kalsnes, B. (2018). Fake news, Oxford Research Encyclopaedia of Communication, Oxford University Press, Vol. 26, 2018, p.1.

Karlova, NA. & Fisher, KE. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. Information Research. 18.

Klonick, K. (2017). The new governors: the people, rules, and processes governing online speech. Harvard Law Review, 6, p.1598-1654

Lasswell, HD. (1927). The theory of political propaganda. American Political Science Review, 21(3), 627–631. Cambridge University Press.

Lewandowsky, S. et al. (2012). Misinformation and its correction continued influence and successful debiasing. Psychological Science in the Public Interest, 13: 106-131.

MacKinnon, R., & Bar, A. (2015). Fostering freedom online: the role of internet intermediaries, UNESCO Center for Global Communication Studies, Paris, 2015, p. 39.

Marwick, A. and Lewis, R., (2017). Media manipulation and disinformation online, Data & Society Research Institute.

Mendel, T. (2015). The UN special rapporteur on freedom of opinion and expression: progressive development of international standards relating to freedom of expression. In T. McGonagle & Y. Donders (Eds.), The United Nations and Freedom of Expression and Information: Critical Perspectives (pp. 235–268). chapter, Cambridge: Cambridge University Press.

Mchangama, J. and Fiss, J. (2019). The digital berlin wall: how Germany (accidentally) created a prototype for global online censorship, Justitia Publisher, p.3-17

Murray, D. (2009). Freedom of expression, counterterrorism and the internet in light of the UK terrorist act 2006 and the jurisprudence of the european court of human rights. Netherlands Quarterly of Human Rights, 27(3), 331–360. https://doi.org/10.1177/016934410902700302

National Commission on Terrorist Attacks Upon the United States (2004). The 9/11 commission report, US Government Printing Office, ISBN 0-16-072304-3

Nowak M., (2005). U.N. Covenant on Civil and Political Rights: Committee Commentary NP Engel, 462.

Oleksy, T. et al. (2021). Content matters. Different predictors and social consequences of general and government-related conspiracy theories on COVID-19. Pers Individ Dif. 2021

Jan 1;168:110289. doi: 10.1016/j.paid.2020.110289. Epub 2020 Jul 22. PMID: 32834288; PMCID: PMC7375289.

Perrone, R. (2014). Public morals and the european convention on human Rights. Israel Law Review, 47(3), 361–378. Cambridge University Press.

Pogácsás, A., Koltay A. (2020). New media and freedom of expression: rethinking the constitutional foundations of the public sphere, Hungarian Yearbook of International Law and European Law (2666-2701): 8, pp. 425-428.

Qualter, TH. (1980). Propaganda theory and the development of propaganda. Ankara University Journal of Faculty of Political Sciences, 1(35). doi: 10.1501/SBFder_0000001406

Redeker, D., Gasser, U. and Lex, G. (2018). Towards a digital constitutionalism? Mapping attempts to craft an internet bill of rights, International Communication Gazette, Cover: 8, p.. 303-325

Rodley, NS. (1993). The universal declaration of human rights: a commentary. Edited by Asbjørn Eide, Gudmundur Alfredson [Oxford/Oslo: Oxford University Press/Scandinavian University Press 1992. 474 pp. ISBN 82-00-21339-0]. International and Comparative Law Quarterly, 42(1), 201–201. Cambridge University Press.

Samples J. (2019). Why the government should not regulate content moderation of social media, Cato Institute Policy Analysis, No. 865. 2019, p. 4.

Sanders, M. L. (1975). Wellington house and British propaganda during the first world war. The Historical Journal, 18(1), 119–146.

Sander, B. (2020). Freedom of expression in the age of online platforms: the promise and pitfalls of a human rights-based approach to content moderation. Fordham International Law Journal, Vol. 43, No. 4, 2020, p. 930-962. http://dx.doi.org/10.2139/ssrn.3434972

Sander, B. (2021). Democratic disruption in the age of social media: between marketized and structural conceptions of human rights Law. European Journal of International Law. Cover 22, Chapter, 32, p. 159-193. 10.1093/ejil/chab022.

Spano, R. (2017). Intermediary liability for online user comments under the european convention on human rights, Human Rights Law Review, Volume 17, Issue 4, Pages 665–679, https://doi.org/10.1093/hrlr/ngx001

Spielmann, D. (2012). Allowing the right margin: the european court of human rights and the national margin of appreciation doctrine: waiver or subsidiarity of european review?, Cambridge Yearbook of European Legal Studies, Cover 14, 2012, p. 381-418.

Sunstein, CR. (2017). #Republic: divided democracy in the age of social media, Ethic Theory Moral Prac 20, 1091–1093 https://doi.org/10.1007/s10677-017-9839-5

Taylor, PM. (2013). Munitions of the mind, a history of propaganda from the ancient world to the present era, Manchester University Press 3rd ed. (July 2013) 7–9.

Taylor, E. (2016). The privatization of human rights: illusions of consent, automation and neutrality, Global Commission on Internet Governance, Paper Series, p. 3.

Tucker, JA. et al. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature, Social Science Research Network, 2018, p. 4. http://dx.doi.org/10.2139/ssrn.3144139

Workman, M. (2018). An empirical study of social media exchanges about a controversial topic: confirmation bias and participant characteristics", The Journal of social media in Society, Vol. 7, No. 1, 2018, p. 382.

Zingales, N. (2019). Intermediary liability in Africa: looking back, moving forward?, Oxford Handbook on Intermediary Liability, p.1-42

Zollmann, F. (2019). Bringing propaganda back into news media studies, Newcastle University, Critical Sociology 2019, Vol. 45(3)

3. Court Practices

Akdas v. Turkey, EctHR, No. 41056/04, 2010, para 27.

Altug Taner Akcam v. Turkey, EctHR, No. 27520/07, 2012.

Aydin Tatlav v. Türkiye, EctHR, No. 50692/99, 2006, para 21.

Chauvy v. France No. 64915/01 2004.

Chorherr v. Austria, EctHR, No. 13308/87, 1993.

Delfi AS v Estonia, ECtHR, No.64669/09, 2015

Erbakan v. Turkey, EctHR, No. 59405/00, 2006, para 46.

Feldek v. Slovakia, EctHR, No. 29032/95, 2001. 137

Gunduz v. Turkey, EctHR, No. 35071/97, 2012, para 28.

Handyside v. United Kingdom, EctHR, No. 5493/72, 1976.

Incal v. Turkey, EctHR, No. 41/1997/825/1031, 1998.

Khurshid Mustafa ve Tarzibachi v. Sweden, EctHR, No. 23883/06, 2008.

Kivenmaa v. Finland, HRC, (Comm. No. 412/1990, U.N. Doc. CCPR/C/50/D/412/1990)

Lesnik v. Slovakia, EctHR, No. 35640/97, 2003.

Lingens v Austria, EctHR, No. 9815/82, 1986.

Mudur Duman v Turkey, EctHR, 15450/03, 2016.

Mukong v Cameroon, HRC, (Comm. No. 458/1991, UN Doc CCPR/C/51/D/458/1991), 1995.

Muller and Others v. Switzerland, EctHR, No.10737/84, 1988. para 35

New York Times Co. v. United States, SCOTUS, (403 US 713), 1971.

Open Door and Dublin Well Woma v. Ireland, EctHR, (App.No. 14235/88), 1992, para 68

Radio France v. France, EctHR, No. 53984/00, 2004.

Semir Guzel v. Turkey, EctHR, No. 29483/09, 2016.

Steel and Others v. United Kingdom, EctHR, No. 851/1058, 1998.

Stevens v. United Kingdom, No. 11674/85, Decisions and Reports Vol. 46

Sunday Times v. the United Kingdom, EctHR, No. 6538/74, 1979.

Sunday Times v. the UK, No. 13166/87, 1991, para 50.

Urper and Others v. Turkey, EctHR, (App. No's. 14526/07, 14747/07, 15022/07, 15737/07, 36137/07, 47245/07, 50371/07, 50372/07 and 54637/07), 2010

Zana v. Turkey, EctHR, No. 69/1996/688/880

Ziembiński v. Poland (No. 2), ECtHR, (Application no. 1799/07), 5 July 2016

4. Other Sources

Statista Statistics of Daily Number of Mails Worldwide: https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/

International Telecommunication Union (ITU), Statistics about internet use, 2022: https://www.itu.int/itu-d/reports/statistics/2022/11/24/ff22-internet-use/

Report of the independent High level Group (HLEG) on fake news and online disinformation, a multi-dimensional approach to disinformation, , March 2018, https://ec.europa.eu/digital-singlemarket/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

Media Or Tech Company? Facebook's Profile Is Blurry' (NPR.org) https://www.npr.org/2018/04/11/601560213/media-or-tech-company-facebooks-profile-is-blurry

Reuters: Facebook CEO says group will not become a media company by Giulia Segreti, 2016 https://www.reuters.com/article/us-facebook-zuckerberg-idUSKCN1141WN

Twitter Removal Requests, Jul-Dec 2021, Republic of Turkey. https://transparency.twitter.com/en/reports/removal-requests.html#2021-jul-dec

Article 19, 'Internet Intermediaries: Dilemma of Liability' https://www.article19.org/wp-content/uploads/2018/02/Intermediaries_ENGLISH.pdf

European Commission (2018). EU Code of Practice on Disinformation, https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

Sounding Board (2018). Unanimous Final Opinion on the So-Called Code of Practice, https://www.euractiv.com/wp-content/uploads/sites/2/2018/10/3OpinionoftheSoundingboard-1.pdf

World Federation of Advertisers (2022). GARM announces guidelines on misinformation, standards on ad placements, and expansion to cover the metaverse https://wfanet.org/knowledge/item/2022/06/21/GARM-announces-guidelines-on-misinformation-standards-on-ad-placements-and-expansion-to-cover-the-metaverse

UK Interim Report: “Disinformation and ‘fake news’: Interim Report Contents, “Russian influence in political campaigns” UK interim https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/36308.htm#_idTextAnchor033

World Health Organisation. “Vaccine misinformation: statement by WHO Director General on Facebook and Instagram”, 4 September 2019, Geneva: https://www.who.int/news-room/detail/04-09-2019-vaccine-misinformation-statement-by-whodirector-general-on-facebook-and-instagram

BBC, “I was a Macedonian fake news writer”, Simon Oxenham, 9th May 2019: https://www.bbc.com/future/article/20190528-i-was-a-macedonian-fake-news-writer

Treisman, Daniel, Foreign Affairs, “Why Putin took Crimea – The Gambler in the Kremlin”, May/June 2016, https://www.foreignaffairs.com/articles/ukraine/2016-04-18/why-russian-president-putin-took-crimea-from-ukraine

SUMMARY

Freedom from Disinformation as an Emerging Human Right

Ugur Petcu

This study reveals the approaches to the notion of disinformation from a human rights perspective and focuses on national and international policies to define, monitor and respond. It aims to emphasise the growing necessity for a universal definition and concept by evaluating regulative responses regarding the spread of disinformation through a human rights perspective.

It provides analysis in the first chapter of the attributes and the spread of false information by comparing different terminologies and its historical uses to properly reflect the phenomenon in order to propose an inclusive terminology. Although this action is evaluated under different classifications today as it was in the past, when it is examined considering its foundation, causes and effects, a definition under the name of disinformation reveals a general concept. It also discloses the provisions that can shape or limit the regulatory responses against disinformation-based activities within the scope of the rights and freedoms offered in the international human rights standards. Scope of protection defined within freedom of expression does not change depending on whether the information shared is true or false. Although this situation largely prevents the limitation against disinformation, it provides certain freedom of movement against serious threats. The last chapter evaluates varying regulative responses of national/international administrations and social media platforms against the spread of disinformation in relation with their effectiveness and applicability. With their effectiveness taken into consideration, penal sanctions provide prevention and deterrence factor, while bringing the possibility of arbitrary interpretations. Therefore, once again demonstrating the importance of reaching consensus on a universal definition of disinformation in terms of preventing arbitrary practice.