#

**SOpHiSM:
*Α RESPONSE TO ONLINE HATE SPEECH THROUGH THE ENHANCEMENT OF HIGH-QUALITY PROFESSIONAL AND CITIZEN JOURNALISM***

***D3.1 Educational Material***

|  |  |  |
| --- | --- | --- |
| **Work Package:** | WP3

|  |
| --- |
| **Capacity Building activities addressed to professional journalists both students and junior journalists, as well as citizen journalists** |

 |
| **Writers:** | Stamatis Poulakidakos and Stylianos Papathanassopoulos |
| **Date:** | 01/12/21 |
| **Version:** | Final |
| **Type:** | Public Document |
| **Place:** | Athens |

|  |  |
| --- | --- |
| Funded by the European Union Rights, Equality and Citizenship Programme (2014-2020) |  |

|  |  |
| --- | --- |
| **Grant Agreement No:** | 875422 |
| **Acronym:** | SOpHiSM |
| **Title:** | A response to online hate speech through the enhancement of high-quality professional and citizen journalism |
| **Start Date:** | 01/03/2020 |
| **Duration:** | 20 months |

**Partners**

|  |  |  |
| --- | --- | --- |
|  | CENTRE FOR EUROPEAN CONSTITUTIONAL LAW (CECL), Coordinator  | Greece  |
| Εικόνα που περιέχει καθιστός, ρολόι, πορτοκαλί, παρακολούθηση  Περιγραφή που δημιουργήθηκε αυτόματα | Hellenic League for Human Rights(HLHR) | Greece |
|  | University of Nicosia (UNIC) | Cyprus |
|  | National and Kapodistrian University of Athens (NKUA) | Greece |

 This document was funded by the European Union’s Rights Equality and Citizenship Programme (2014-2020). The content of this document represents the views of the authors only is their sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

# Introduction

# Aim and structure of the seminar “Identifying and tackling online hate discourse”

As the title suggests, the main purpose of these seminars is, on the one hand, to familiarize participants with the different definitions, interpretations and approaches regarding (online) hate discourse and, on the other hand, to provide them with the necessary ‘tools’ to identify and ‘combat’ hate discourse online and/or offline.

Due to the multiple cultural parameters that permeate the structure and manifestation of hate discourse, this educational material does not aim to provide an ‘exhaustive’ overview of this social phenomenon, but instead represents a first attempt at a coherent approach in light of an ideologically ‘fruitful’ discussion that is used publicly with specific purposes and consequences, both on an individual and collective/societal level.

# 1.1.1 In the beginning, there was Discourse

It is precisely for this reason that we contend that the Greek rendering of this phenomenon be illustrated with the term “*Λόγος (with a capital ‘Λ’) μίσους” [/ló.ɡos/ /mí.su:s/],* which corresponds to the English *hate discourse* (and not *hate speech*, whose accurate Greek translation is *“ομιλία μίσους”* *[/o.mi.lí.a/ /mí.su:s/]*, though it is often translated as *“ρητορική μίσους”* [*/ɾi.to.ɾiˈci/ /mísu:s/]),* as the term *“Λόγος” primarily refers to the strongly ideological and socially constructed dimension of expressed hatred and, also, conceptually includes the expression of hatred through other forms of communication besides speech or written text, such as animated and still images or a combination of written text and images (e.g. illustrated cartoons)*. Moreover, the term “hate discourse” (*“Λόγος μίσους”)* highlights the distinct social dimension of this phenomenon, in contrast to the term “hate rhetoric” *(“ρητορική μίσους”)* which focuses on the expression of hate through language (either speech or writing) as a verbal/rhetorical scheme. Of course, we must note at this point that in the broader public debate, the term “hate rhetoric” has prevailed, possibly because it is more phonetically pleasing, and thus we also make use of it on our website. However, in this manual we ought to properly explain the term and provide a correct basis for our analysis.

In the next section, we lay the groundwork for reading and understanding this handbook (and discussing it during the seminars) by introducing the example of Social Constructionism, which focuses on the meaning of *Discourse* *(“Λόγος”)* and the ideology it represents as the cornerstone of the construction and interpretation of social reality.

In the second section, we attempt a very brief overview of the Internet as a space of open communication, the main features that define it, and the fact that it constantly becomes a space for the expression of hate discourse.

Then we focus on the concept of hate discourse, examining different approaches and discussing the causes of its manifestation and its possible consequences at the individual and societal level.

Next, we look at (online) hate discourse beginning with the debate on how to legally address the phenomenon.

Finally, we focus on strategies to combat hate discourse that focus on providing education, both online, in the sense of digital ‘training’, and offline, on a more social level, to create socially intelligent and digitally literate individuals, while suggesting some ways to avoid expressing and participating in hate discourse.

# Social constructionism

*Social constructionism* describes the approach that accepts that the relationship between the individual and society, as well as the relationship between humans and nature, is one mediated by discourse (Demertzis 2002:141). As an organized set of relations, society is a construction insofar as its members, knowing each other, recognize it as the framework of their common life. This (mutual) recognition is realized through discourse, the set of semiotic codes (language, symbols, etc.) through whose rational and intentional use the subject shapes its sense of self and the surrounding world. In other words, *discourse is the main path of social interaction,* it structures the experience of reality and human meanings, it ‘acts’, it has functions, it embodies ideology and purpose (e.g. it demands, justifies, accuses, etc.), while various rhetorical techniques are used to achieve its goals (Thompson 1999:7).

 According to this logic, the meaning individuals attach to their actions is neither arbitrary, random, nor timeless. The *social construction of reality* is realized through the “externalization - objectification – internalization” dialectic. Human externalizations (creations, practices, artifacts, ideas, symbolisms), in order to acquire semantic substance, encounter the experience of previous generations and are simultaneously inscribed in the historical dynamics of the experienced culture. Thus they are ‘objectified’, transformed into ‘reality’, into ‘real’, ‘natural’ elements of the world. Both the subjects themselves and the next generations internalize them as self-evident facts of the social system and the experienced world. In this way, the social construction of reality, i.e. the way we understand and construct ourselves, society and nature, is a continuous and ‘dynamic’ process(Demertzis 2002:141-142).

 The social construction of reality can be distinguished into two -interconnected - levels: *Construction* and *Manufacture.*

 ***Construction*** is the totality of institutional events that make up the Social and the Political, the ontologically given ‘subjective’ reality. At the same time, it is (for the individual) the ontologically ‘objective’ reality. It is a structured system of values, institutions, and social rules that we (as individuals) face in society and encounter through our socialization in the social context into which we are born and in which we develop as personalities. The political system, for example, is an institutional event; it ‘exists’ only through shared conventions, goals, and meanings. An earthquake, on the other hand, is an objective event that does not depend on moods, contracts, preferences, or choices. Yet it is symbolically valued, assigned social meaning, and scientifically studied through the general system of language, representations, and cognitive forms.

 ***Manufacture*** presupposes Construction. It is the totality of intellectual creations, imaginary inventions, delusions, and illusions that are ‘subjective’ from a scientific point of view and represent unfounded assertions or evaluations. They are the ways in which we evaluate, ‘reproduce’ and ‘metabolize’ the social parameters of Construction, i.e. the institutions and conventions that exist and constitute society, as well as the individuals with whom we live and work within a particular social framework (Demertzis 2002:175-178). In other words, the individuals themselves as social subjects constitute social constructions (Miskolci, Kovacova, Rigova 2018:2). It is at this level of Manufacture that we find hate discourse, which negatively ‘evaluates’ social institutions and conventions as well as individuals or groups of individuals on the basis of subjective (ideological) criteria, creating and at the same time reproducing a framework of inequality and discrimination. We will examine this specific form of discourse -primarily- within the communication environment of the Internet, following the logic of the Discourse Analysis approach.

 Social discrimination is a widely studied topic in Discourse Analysis, which explores the importance of language in establishing, maintaining, resisting and changing social power relations, mainly through the ideological functions of political and media discourse (Blommaert and Bulcaen 2000).

*Through its repetition, discourse that reinforces social discrimination ‘constructs’ assumptions, legitimizes domination and naturalizes inequality (Assimakopoulos, Baiden and Millar 2017:5)*.

 At the same time, Critical Discourse Analysis focuses on the relationship between ideology, inequality, and power through discourse, analyzing these components on the basis of “opaque as well as transparent structural relationships of domination, discrimination, power, and control as manifested in language” (Wodak 1995:204).

 One of the basic principles of Critical Discourse Analysis is that **social interaction (partly) takes a linguistic form**. Therefore, the focus is on linguistic features and the organization of certain meanings in discourse. For example, the use of a passive voice in news reports about the deportation of immigrants or an attack on a transgender person may result in the concealment of the perpetrator(s) and thus minimize the possibility of accountability (Assimakopoulos, Baiden and Millar 2017:5).

# The online communication environment

Towards the end of the 20th century the rapidly expanding technological revolution brought information technologies to the fore (Castells 2000:1). The exchange of messages between people is now subject to far fewer spatial and temporal constraints (Poster 1995:18; Mansell and Steinmueller 2002), as communication content is now instantly and globally available (McNair 2006:4-5) and can be stored and accessed from anywhere as long as there is electricity and suitable equipment.

The Internet is the epitome of this revolution in communication. Its high adaptability and flexibility of use make it an invaluable tool in the hands of those who wish to get their message across. The Internet has created an entirely new structure of communication relations, a system of multiple producers, distributors and consumers, where the boundaries between these concepts ‘collapse’ (Poster 1995:3, 28; De Latour et al. 2017:38).

The continued expansion of the Internet since the 1990s has had a significant global impact. We have moved from an information transmission system dominated by traditional one-way media, the state and powerful political and economic lobbies, to an “information society” (Pleios 2011:206) in which citizens are not merely consumers and (re)transmitters of information, but can also take a more active role as creators and co-creators of new content, the so-called “pro-sumers”. It is therefore hard to deny that the rapid expansion of the Internet has affected and continues to affect societies on a micro, medium and macro-scale (Assimakopoulos, Baiden and Millar 2017:10).

The rapid growth of the Internet is due to a number of factors that have greatly influenced the field of communication, more specifically:

the mediation of an ever-increasing part of social experiences and relations, i.e. the formation of modern societies as media societies (Schmidt 2001:11) and

1. the fusion of different forms of messages (written, audio, audio-visual, visual) within the Internet communication ‘space’ and the possibility of sending a large number of messages with different content (Webster 1995; Terranova 2004:20-27).

The changes in communication brought about by the Internet (e.g. interactivity, search engines, social media, production and dissemination of a large amount of information) represent a continuation and, at the same time, a shift in the previous -unidirectional- nature of communication and the media (press, television, radio, cinema) that characterize it (Pleios 2011:144; Jakubowicz et al. 2017:6).

The fact that the Internet offers users such a wide variety of services has favored its rapid growth and tranformed it -after thirty years of commercial use- into a very important tool for billions of people. At the same time, however, the Internet has been 'accused' of representing a 'virtual' globalization -and the associated techno-cultural imperialism- of the planet, as web content appears to be "centered" around specific applications such as Google and Facebook, and content from certain countries such as the US, China, Russia and Japan (Jakubowicz et al. 2017:6). This shows that the Internet is indeed not a 'closed' electronic space, but is intimately connected to the social environment, with an intensity that fluctuates, spreads and varies according to the rhythm of geopolitical events, social issues and cultural trends that form the framework that surrounds it (Blakemore and Longhorn 2001; Wiggins 2001; Jakubowicz et al. 2017:1). In short, ***the Internet reproduces -among other things- the power relations that also manifest in offline social life***.

The Internet has changed the way people -especially young people- interact, both interpersonally and socially (Oksanen et al. 2014:269). Almost every kind of interest and activity is represented by an online community, as the Internet can accommodate almost anything: from the most specialized scientific research to the most insignificant hobbies, from the greatest religions to the worst human perversions. The conflicting views that prevail among theorists who have studied the Internet are vividly illustrated by the words of Noam Chomsky (2003:16, 51), who recognizes positive elements in the functioning and innovations of the Internet, but at the same time points to certain areas of concern.

The Internet has contributed, on the one hand, to the strengthening of relations between individuals and social groups and, on the other, to the emergence of spaces and tools of hatred.

On the one hand, the “tools” offered by the Internet have contributed to the strengthening of relations between individuals and (social) groups (e.g., the Internet has made it possible to organize social movements worldwide and has facilitated cross-border cooperation) (Poulakidakos, Veneti, Frangonikolopoulos 2018). On the other hand, however, ***the Internet and social networking platforms in particular are also used as 'spaces' and tools of hate***. Although there is no statistical data providing a comprehensive overview of this phenomenon, social media platforms and organizations established to combat hate discourse claim that online hate discourse is an increasingly common phenomenon (De Latour et al. 2017:38; Jakubowicz et al. 2017:55).

Communication, including the production and exchange of information, is one of the main features of the Internet. However, this type of digital communication is characterized by certain peculiarities: ***The Internet is a space that allows users to express their views and communicate without boundaries and usually (though not always) without control***. As was also noted by interviewees on the topic of online hate discourse, “*anonymity plays a very big part in expressing hate on the Internet, there is a freedom that allows you to express your worst self*” (YJ)[[1]](#footnote-1). The ‘grammar’ of the Internet makes it easier for users to believe that they are hiding their identity (in whole or in part) and in some cases even their location and activities: “*Hate rhetoric can be easily expressed on the Internet because it allows [users] to hide behind anonymity, so they feel free to say whatever they want without taking personal responsibility for a comment or text*” (YJ). However, although most of the public communication on the Internet can potentially be traced back to its origin, most users view the Internet as a platform that allows them to express themselves freely and anonymously (Assimakopoulos, Baiden and Millar 2017:10-11).

Since the Internet is a socio-technological infrastructure, more precisely a communication space that -from a technological point of view- offers new possibilities to satisfy our communication needs, but at the same time offers a relatively "open" architecture whose use is (co-)shaped in relation to the social context in which it exists and the communication habits of its users (Mollen 2020:227), it can -among other things- become a space for the generation of hate discourse.

The communication possibilities offered by the Internet have been exploited for the expression of hate (Jakubowicz et al. 2017:1). From its earliest beginnings, the emergence of the Internet and its use as a public communication tool has been exploited by individuals and organizations with a racist agenda to generate and disseminate hate discourse (Jakubowicz et al. 2017:2). This expression of hate discourse on the Internet cannot be ignored as it affects the majority of young people for whom the internet is an integral part of their daily lives (Oksanen et al. 2014:270).

***Because of its global, direct, and participatory nature, the Internet has become a place where intolerant ideas and beliefs are expressed and spread,*** providing an additional means of disseminating discrimination based on prejudice, which can potentially even lead to hate crimes: “*The targets of hate rhetoric are refugees, immigrants, people with different sexual orientations, women, animal lovers*” (YJ). Such behaviors reject diversity and aim to strip individuals and groups of their dignity by attacking their identity (Assimakopoulos, Baiden and Millar 2017:11). The spread of hate discourse on the Internet can have a devastating impact on the social fabric, negatively affecting both targeted groups and/or individuals and those who advocate for freedom, tolerance and non-discrimination, while also negatively affecting the dialogue that takes place on online platforms (European Commission 2016b:1).

***Examples of hate discourse expression can be found in most countries around the world and show that intolerance and hatred can be developed on the Internet,*** taking advantage of its very nature (Gagliardone et al. 2015). Although the ‘terms of use’ of most social media platforms, such as Facebook, Yahoo!, or Twitter, prohibit the posting of "illegal, harmful, defamatory, obscene, vulgar, invasive, or racially, nationally, or otherwise hateful" content (Cohen-Almagor 2015:163), the time usually required to remove such content is a matter of growing concern. This has led the European Commission and various social media technology giants to agree on a Code of Conduct, specifically targeting illegal Internet hate discourse (European Commission 2016b; Assimakopoulos, Baiden and Millar 2017:11-12).

It is particularly difficult (if not impossible) to control all the possible ‘dangers’ that arise from the use of the Internet. In fact, given the connection between the online environment and offline reality, it may be easier to reduce online risks by effectively addressing a variety of offline factors; indicatively, and according to relevant youth research, cyberbullying is associated with an insufficient preoccupation with school/student homework and loose to non-existent family ties. At the same time, the psychosocial problems that young people face in their offline life overlap with their negative online experiences (Oksanen et al. 2014:270).

Additionally, we need to talk openly about the intense manifestation of hostile online behavior. For example, in recent years, the scale and intensity of anti-feminist sentiments on the Internet have become a major concern, not only for feminist activists but also for any woman who openly expresses her views in online forums. An Amnesty International report published in 2017 showed that almost a quarter (23%) of women surveyed across eight countries stated they had been cyber-abused or harassed at least once (Ging and Siapera 2019:1). The picture is similar regarding the manifestation of racial, racist hate discourse (Jakubowicz et al. 2017).

Based on this logic, the next section focuses on definitions/approaches to (online) hate discourse, its particular characteristics, its causes, and consequences.

# Summary

In the first section of this educational material, we lay the ‘bases’ for approaching the phenomenon of hate discourse. First, due to the profoundly social nature of the phenomenon, we propose the term “hate discourse” instead of “hate rhetoric” (or “hate speech”) and, hence, we trace hate discourse in the realm of social Manufacture, i.e. in the negative subjective (ideological) evaluation of institutions and conventions of society, as well as of individuals or groups of individuals creating and reproducing a framework of inequality and discrimination. In this light, hate discourse can be approached through (critical) discourse analysis, which emphasizes the process of constructing social reality through discourse.

In parallel, we describe the communication environment of the Internet, which is at the center of the interest for this program, as a ‘space’ through and within which -unfortunately- numerous cases of hate discourse take place systematically.

# Hate discourse

# Stereotypes - prejudices - narratives

According to the social constructionism approach, the representations of individuals, groups, events, issues, phenomena, and relations through discourse are both formed by and form the socio-political status quo of these entities (Assimakopoulos, Baiden and Millar 2017:54). As the stereotypes and prejudices that shape the socio-political status quo of individuals and social groups are constructed and disseminated through discourse, we will begin the review of hate discourse from these two concepts. Then, we will briefly refer to the narratives that are -as the products of discourse- interpretations of the social reality that may contain and diffuse prejudices, and we will conclude with an overview of the different definitions and approaches regarding hate discourse, in order to formulate a definition that will correspond to the logic by which the phenomenon is approached in the present educational material.

The notion of stereotype is closely linked to prejudice, to the extent that the two concepts are often confused. We could argue that a stereotype is the potentially cognitive core of prejudice, a set of information and beliefs relating to a particular category of subjects (e.g. social groups) that have been transformed into a unique, coherent, stable image capable of maintaining and creating prejudice perceptions against them (Assimakopoulos, Baiden and Millar 2017:26). In other words, the stereotype signifies the mental process of classifying reality based on particular information/characteristics, while prejudice is the emotional projection of (negative) value judgments on these entrenched mental categorizations. Despite their close relationship, it should be noted that the stereotypical categorizations of the complexity of the world around us do not necessarily and unavoidably turn into prejudices.

In terms of creating prejudices, the process of (stereotypical) generalization creates a mechanism through which the individual tends to group all individuals presenting characteristics considered incompatible with their own -ideologically charged- view of the world (Weltanschauung). Through this process, which is generally known as “Othering”, a social group is considered not to belong to the individual's group, in terms of identity characteristics such as gender, nationality, religion, etc. This categorization is often accompanied by a negative -based on qualitative criteria- categorization of Others and a consequent denial of recognition of fundamental human rights (e.g. free expression, dignified living) (Assimakopoulos, Baiden and Millar 2017:26).

Social groups that have been ‘constructed’ by an individual as different from their own, based on their skin color, sexual orientation, mother tongue (Miskolci, Kovacova, Rigova 2018:2) are often categorized as inherently ‘inferior’ as if their members were born with a specific set of threatening characteristics. Filth, odor, criminal behavior, and immorality are just a few examples of characteristics and behaviors attributed to these (different) social groups as general, inherent, and unchanging features. The processes of creating and reproducing these social constructions take place in all possible environments, through the mainstream media, social media, art or science, or even daily interactions with colleagues, relatives, friends, etc. In this way, the very existence of messages about the way certain people ‘look’ and in what ways they behave make these characterizations seem true, factual, and indisputable (Foucault 1977, 1979; Miskolci, Kovacova, Rigova 2018:3).

The repetition and high frequency with which hate discourse can take place in society create a framework of widespread social acceptance of the view that certain social groups and/or minorities are indeed inferior and should be treated accordingly. In such cases, social media, such as Facebook and Twitter, play a key role as communication channels (Miskolci, Kovacova, Rigova 2018:2). A typical example of this condition is the predominantly negative public discourse regarding refugees/immigrants in Greece (Poulakidakos 2018), which, of course, has also spread on the Internet.

The repetition and high frequency with which hate discourse can take place within the context of society creates a framework of widespread social acceptance of the view that certain social groups and/or minorities are indeed inferior and should be treated accordingly.

This condition becomes even more pronounced when ‘defensive instincts’ are activated in response to the perception of Others as a threat. The **devaluation of the Other** takes mainly the form of verbal hostility against different cultures, while, in many cases, people ‘disguise’ their aggression as a form of defense, emphasizing the ‘dangers’ posed by a national or religious group (Taguieff 1993; Gale 2004). Hate discourse is based on tensions which it seeks to reproduce and reinforce; for example, tensions that escalate during political crises and lead to polarization (Poulakidakos and Veneti 2016), such as during the economic crisis or the discussion regarding the management of the immigration issue ( Mollen 2018:236).

Such communication acts can, at the same time, ‘unite’ and ‘separate’ people, creating ‘us’ against the ‘others’ (Gagliardone et al. 2015:11), the ‘good’ and the ‘bad’, in a Manichaean-logic distinction of the members of one or more social groups. In this light, hate discourse can also take on a de facto propaganda function (Poulakidakos 2014): “*There are specific interests behind hate discourse. For example, certain online media promote the Golden Dawn [political party]. Hate discourse is used as a means of propaganda*” (FGHR). Therefore, hate discourse can be disseminated directly or indirectly by politicians or other public figures on the Internet, ultimately rendering it ‘acceptable’ in the eyes of the public. Politicians can use prejudices to provoke unreasonable public fears and win the popular vote (Miskolci, Kovacova, Rigova 2018:1): “*Stereotypes, prejudices and the creation of scapegoats during a crisis (e.g. [the example of] Golden Dawn in Greece) contribute to the spread of hate discourse in the public sphere*”(YJ).

Creating and reproducing stereotypes is an ongoing process, as stereotypes are constantly evolving. History offers numerous examples of initially neutral stereotypes (Assimakopoulos, Baiden and Millar 2017:29) that gradually turned into prejudices, involving initially isolated and later collective manifestations of hate discourse and, finally, actions and crimes, including harassment or even the mass extermination of vulnerable social groups, such as the Romas, Jews, Armenians, homosexuals, etc. (Assimakopoulos, Baiden and Millar 2017:30).

The timeless consolidation of specific stereotypes and prejudices, i.e. interpretations of social reality (Manufacture), can reinforce these interpretations, ‘solidify’ them and ‘transform’ them into (pseudo-)Constructions, which are spread through discourse from generation to generation and, thus, acquire timelessness.

Prejudices are quite difficult to change. Prejudice usually resists the ‘test of reality’, as conspiracy theories clearly show. For example, for those who deny that man ever went on the moon, the photos of Neil Armstrong walking on the moon could have been manufactured, as there is no picture of the person taking the photo. The ‘movement’ challenging the existence or severity of the coronavirus follows a similar logic, as its members dispute the existence of the epidemic, resist the use of masks as a preventive measure -despite the scientific data- and despite the repeated recommendations of experts refuse to follow the relevant instructions.

However, prejudice cannot be fought by prejudice. It must be offset or balanced by events and strategies that invite and motivate people to see other aspects of the same reality, hoping to interpret them differently (Del Felice et al. 2017:14).

When a narrative is presented as the only correct one, rejecting alternative narratives or, in extreme cases, inciting violence against those who question it, in other words, taking the form of hate discourse, the fundamental rights of a democratic society are at stake.

Stereotypes and prejudices are spread through discourse in the form of narratives. A narrative is a logical, or at least seemingly logical, and internally coherent interpretation of interconnected events or information that makes sense to the reader/listener (Del Felice et al. 2017:9). When a narrative is presented as the only correct one, rejecting alternative narratives or, in extreme cases, inciting violence against those who question it, in other words, taking the form of hate discourse, the fundamental rights of a pluralistic society (e.g. the right to freedom of thought, freedom of expression, freedom of religion and belief, etc.) are at stake (Del Felice et al. 2017:9). “*In social terms, hate discourse perpetuates the stereotypes that fuel it, creating a vicious cycle*” (YJ)*.*

# 2.2 The issue of defining hate discourse

Hate is a social phenomenon that has existed in human societies since their creation. Throughout the course of human history and the construction of societies, hatred has led to disastrous consequences, such as wars and genocides (Royzman, McCauley, and Rozin 2005; Heinze 2016:2). Hate discourse has undoubtedly led to violence and discrimination in democracies such as Weimar, Germany, and the civil wars in Yugoslavia or Rwanda (Heinze 2016:207).

Since the second half of the 20th century, Western liberal democracies have embraced -at least de jure- the ideals of non-discrimination and equal treatment of all before the law (Heinze 2016:3). Hate discourse is an integral part of a complex chain, along with freedom of expression, individual, social and minority rights, as well as concepts such as dignity, freedom, and equality (Gagliardone et al. 2015:10).

Many definitions of hate discourse have been proposed over the years. In international jurisprudence, hate discourse often refers to expressions that promote incitement to violence (discrimination, hostility, and/or violence), based on the fact that the target is identified as part of a particular social group. This may include -but is not limited to- discourse that supports, threatens, or encourages violent acts.

On a primary approach level, hate discourse is defined as a particularly harmful discourse contributing to the formation of a climate of hatred and violence, especially towards marginalized and weaker social groups, violating the fundamental dignity of its victims (Gelber 2002:1), that is, violating the principle of equal treatment. As we will see later, there are a number of definitions of hate discourse that can be categorized based on the aspect on which they focus most. This has created a universe of definitions that reflect the complexity and consequent difficulty of defining the phenomenon.

Hate discourse functions in many cases as an ‘empty signifier’ (Laclau and Mouffe 1985), since it is a term that may seem obvious to most, but for which people tend to provide quite diverse descriptions (Gagliardone et al. 2015:55; Waldron 2012:27). Despite repeated attempts by different actors to establish a specific definition, hate discourse is still used as a rather concise and broad term (Waldron 2012:34), ranging from direct threats to the safety of individuals and (vulnerable) social groups to cases in which individuals can vent their anger against those in power (Gagliardone et al. 2015:7; Antigone 2019:3; Waldron 2012:27). Therefore, the concept of hate discourse, both offline and online, includes many different parameters related to offensive content (Miskolci, Kovacova, Rigova 2018:2).

In national and international jurisprudence, hate discourse refers to expressions that promote incitement to violence (discrimination, hostility, and/or violence), based on the fact that the target is identified with a particular social group. This may include -but is not limited to- discourse that supports, threatens, or encourages violent acts. For some legislations, the concept extends to expressions that create a climate of prejudice and intolerance, as this, too, can trigger targeted discrimination, hostility, and violent attacks (Gagliardone et al. 2015:10). In several countries, lawyers refer to group libel or group defamation (Waldron 2012:39).

In today's digital environment, cyber-mediators -intermediaries for the processing of Internet communication such as Facebook, Twitter, and Google- have proposed their own definitions of hate discourse that bind users to a framework of rules and allow the restriction of certain forms of expression. National and regional actors are also trying to promote a logical approach to defining the term that is more ‘grounded’ in local social norms. Within this framework, however, the possibility of achieving a universally accepted definition seems unlikely (Gagliardone et al. 2015:7-8; Antigone 2019:3).

In today's digital environment, cyber-mediators such as Facebook, Twitter, and Google, have put forward their own definitions of hate discourse that bind users to a framework of rules and allow the restriction of certain forms of expression.

In everyday life, ***"hate discourse" has become an "umbrella" term with multiple meanings that covers a heterogeneous collection of communication aspects*** (Brown 2020:47). This heterogeneity is reflected in the description of hate discourse provided by the European Commission's General Policy Recommendation against Racism and Intolerance (ECRI GPR 2016), which:

“Considers that hate speech is to be understood as the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of “race”, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status. Recognises that hate speech may take the form of the public denial, trivialisation, justification or condonation of crimes of genocide, crimes against humanity or war crimes which have been found by courts to have occurred, and of the glorification of persons convicted for having committed such crimes; Recognises that the use of hate speech may be intended to incite, or reasonably expected to have the effect of inciting others to commit, acts of violence, intimidation, hostility or discrimination against those who are targeted by it and that this is an especially serious form of such speech”

At the same time, as we are constantly confronted with discreet and divisive attitudes and actions in our society, the word “hate” appears more and more often in public political discourse, in order to serve ideological/political expediencies. ***The fact that there are so many forms of hate discourse should not surprise us***, as it serves many different purposes, some of which are intertwined with intense political controversy (Brown and Sinclair 2019: ch. 1; Brown 2020:48).

In the colloquial language, however, approaches to hate discourse tend to be broader, extending to insults against people in power (e.g. politicians) or celebrities. Especially in politically critical moments, such as during elections, the concept of hate discourse can be used as a means of propaganda to discredit political opponents: accusations of hate discourse incitement can be exchanged between political opponents or used by those in power to reduce public disagreement and criticism (Gagliardone et al. 2015:10).

In many European countries, as well as several other countries around the globe, the term “hate discourse” has become an ‘ideological recrimination’ used by political opponents accusing each other publicly of ‘unacceptable discourse’. In this way, they manage to incite public controversy and silence (political) criticism (Brown 2020:47; Special Rapporteur on the promotion and protection of freedom of opinion and expression 2019:4; Strossen 2018:1-2) Numerous political controversies, and the intense rhetoric they often include, have provoked mutual accusations of “hate discourse”: “*Hate discourse is part of political communication and is used as a means of propaganda. Nowadays, anyone can attack anyone through social media, without any limits*” (YJ). For example, members of the Black Lives Matter movement have been accused of expressing “hate discourse” against police officers, while critics of the Black Lives Matter movement have been accused of expressing “hate discourse” against its supporters or against the African-American community in general (Strossen 2018:11).

At the same time, the term “hate discourse” as an aggressive delineation, accompanies a wide range of controversial public discourse, from false news to anti-abortion demonstrations (Strossen 2018:12). Research has concluded that hate discourse in many western democracies stems from ideological divisions and recriminations between representatives of the left and right ideologies of the political spectrum, which -among other things- poses serious challenges for journalists in their profession (Cochliou 2020).

Besides, and based on what experience teaches us, anyone can be accused of either practicing or ‘accepting’ ‘hate’ based on personal characteristics and beliefs (Strossen 2018:1).

In addition, a typical example of the different and ideologically colored approaches to address hate discourse is “the issue of hate discourse on the Internet”, about which public perceptions are radically divergent and deeply rooted in political disagreements and ideological schisms (Brown and Sinclair 2019: ch. 1).

An initial and relatively simple definition of hate discourse describes it as “an expression that conveys views of hatred or discrimination against specific individuals or groups” (Strossen 2018:1). On a politically-centralized approach, the Council of Europe refers to a term “used to describe broad discourse that is extremely negative and constitutes a threat to social peace”. ***Hate discourse covers all forms of expression*** (speech, writing, images, videos, or any other form of communication online, or offline) (Waldron 2012:5) that spread, incite, promote, or justify racial hatred, xenophobia, antisemitism, or other forms of intolerance-based hate, including intolerance expressed through aggressive nationalism and ethnocentrism, as well as discrimination and hostility against minorities and immigrants (Antigone 2019:2; Del Felice et al. 2017:8; 27).

According to the “No Hate Speech Movement” additional forms of discrimination and prejudice, such as discrimination against the Romas, Christianophobia, Islamophobia, misogyny, sexism, and discrimination based on sexual orientation and identity, all fall within the scope of hate discourse (Del Felice et al. 2017:27).

Although it may lack clear definitions, hate discourse (regardless of the way it is being expressed) can be identified through its derogatory and/or inhuman purposes. When approached as a rhetorical strategy, hate discourse must be understood as “the intentional use of words, images and symbols, hyperlinks, downloads, news, conspiracy theories, political views, and even the content of popular culture to create an effective mechanism of provocative/inflammatory discourse” (Jakubowicz et al. 2017:55).

The strong socio-political causes behind the expression of hate discourse are best illustrated in the approach describing hate discourse as a “manifestation of unequal social and power relations and, at the same time, a mechanism for their reproduction” (Del Felice et al. 2017:30) and connecting it to social norms, which emerge as a result of prejudices.

*Social inequality* is reproduced in different ways, for example through mechanisms that offer different/unequal access to opportunities and (social) resources to different (social) groups, creating privileges for some, and leading others to marginalization. These hierarchical relations between social groups are established based on prevailing social norms. Social norms are rules that define the expected behavior of individuals within a social context (they are part of Construction, as they roughly constitute established principles of operation of any society, they are the focal point of ‘social DNA’). These social norms are the key to securing privileges for individuals or social groups whose daily practices and socially recognized positions are closer to established social norms and ideals. At the same time, the same social norms also define the disadvantaged and marginalized groups. The smaller the deviation from the established social rules and values, the more privileges those individuals and groups enjoy. The greater the deviation, the less the acquired privileges, and the more intense the marginalization of individuals or groups (Del Felice et al. 2017:32).

*Social inequality* is reproduced in different ways, for example through mechanisms that offer different/unequal access to opportunities and (social) resources to different (social) groups, creating privileges for some, and leading others to marginalization.

*Social norms* are reproduced, strengthened and modified at every level of social life, from interpersonal relationships to interactions within institutions and organizations, as well as in media representations of society, both traditional and new. Deviations from social norms are often ‘punished’ through different mechanisms. One of these mechanisms is hate discourse, which seeks to strengthen social norms that can be ‘destabilized’ due to deviations or provocations against them (Del Felice et al. 2017:32).

Hate discourse primarily targets social groups whose position in society is inferior to other groups. In order to deal with hate discourse effectively, it is important to perceive it as a violation of human rights and as a social phenomenon with profound (social) causes.

Hate speech primarily -but not exclusively- targets social groups whose position in society is inferior to other groups (e.g. minorities) (Waldron 2012:27), and/or whose ideas and behavior collide with the predominant system of social rules (Del Felice et al. 2017:33). Hate discourse, therefore, is not a separate characteristic of a particular social group; it arises in the context of specific social (power) relations, occurs when social norms and privileges are challenged and can be expressed ‘directly’ or covertly. In order to deal with hate discourse effectively, it is important to perceive it as a violation of human rights and as a social phenomenon with profound (social) causes (Del Felice et al. 2017:36).

Although anyone can potentially become a victim of hate discourse, in Europe there are particular social groups that are frequently and specifically targeted (Del Felice et al. 2017:33). Various studies have named LGBT+ community members, women, immigrants, ethnic minorities, Romas, Jews, the impoverished, people with disabilities, Muslims, and other religious minorities as frequent targets of hate discourse. “*The social groups that are the most frequent targets of hate discourse are women, Romas, foreigners and people with disabilities*” (YJ).

Unfortunately, there is insufficient data on young people as targets of hate discourse, but certainly hate discourse targets the youth just as much as any other age group, and tends to have more severe consequences, due to the vulnerability of many young people (Del Felice et al. 2017:33): “*Children experience the most severe consequences of hate discourse*” (JS). "*Words can be just as painful as actions. Children have no limits when it comes to making fun of someone, but they cannot assess the exact consequences of their actions, both for themselves and for others*” (CJ). In this context, “Hate speech is considered speech that is particularly harmful, because it contributes to creating a climate of hatred and violence against marginalized and weak sectors of society, fundamentally violating the human dignity of its victims” (Gelber 2002:1).

Our right to freedom of speech must not be turned into aggression, harassment, and threats that result in the obstruction of the right to freedom of speech of others.

For some theorists, a basic dividing line in order for a statement to be considered hate discourse is drawn when one person's discourse threatens another's right to speak and/or attempts to silence them. Our right to freedom of speech must not be turned into aggression, harassment, and threats that result in the obstruction of the right to freedom of speech of others (Porter 2020:xx). On a theoretical level this ‘dividing line’ is relatively easy to describe, but on a practical level it is the subject of intense debate for scholars of law and ethics, the justice system and society as a whole (Porter 2020:xx).

Following a similar logic, an observer could approach hate discourse as “a set of verbal acts with a specific strategic purpose: to insult, humiliate, degrade, create/strengthen a framework of discrimination, hurt, threaten” (Gelber 2002:63).

Hate discourse is, therefore, a form of expression that creates and perpetuates a framework of discrimination (Gelber 2002:69). These areas of discrimination include:

1. the restriction of the victims' individual freedoms,
2. the ‘normalization’ of messages that promote social discrimination, to the extent that the reader/listener/viewer begins to truly believe the allegations of inequality,
3. the manifestation of further acts of submission, and
4. the concealment of opposing views (Gelber 2002:83), thus contributing to the halting of the socialization process and proper development of the victims' personalities (Gelber 2002:86-87), as
5. verbal acts of hatred limit the ability of their immediate -and other potential- targets to respond and challenge hate discourse claims (Gelber 2002:88).

This framework of approaching hate discourse leads to the characterization of verbal acts of hatred as verbal acts of discrimination (Gelber 2002:87).

In this way, hate discourse aims to degrade the dignity of those it targets, primarily in the eyes of members of society. It aims to tarnish their reputation by linking identity characteristics such as nationality, race, gender, social class, and religion with behavioral traits that should exclude one from being treated as an equal member of society (Waldron 2012:5; Sobieraj 2019).

Based on the individual or social characteristics it seeks to discredit, hate discourse can be described as an expression of hatred against an individual or group of individuals on the basis of protected characteristics. The term “protected characteristics” implies identity parameters (biological characteristics, sexual orientation, cultural characteristics) that identify individuals or (social) groups. What these protected characteristics are remains open to different interpretations, with different states incorporating different features into this categorization (Assimakopoulos, Baiden and Millar 2017:3). For example, the definition of the EU (Antigone 2019:2; Del Felice et al. 2017:8; 27) leaves gender, gender identity, and sexual orientation out of the protected characteristics definition.

One of the unique features of the attacks against identity characteristics is that they are both personal and general. Although identity bullying may seem like individual bullying, the anger it harbors is structural, expressing hostility towards the voice and visibility of specific speakers perceived, in this case, as (informal) representatives of specific social groups (Sobieraj 2019).

Another issue that arises even with regard to hate discourse definitions which do not refer to specific “protected characteristics”, like the UN's approach, which defines hate speech as “incitement to hatred, discrimination, hostility, or violence", is the interpretation of the words “hate”, “discrimination”, “violence”, and “hostility”, which is often based on ideological criteria. Nonetheless, the most important aspect of this definition (and others like it) is the word “incitement”, which plays a leading role and signals the intention to provoke possible hostile actions against members of social groups, and acts as a precondition for assessing any form of discourse as hate discourse, suggesting a link between hate discourse and hate crime, with the former leading to the latter (Assimakopoulos, Baiden and Millar 2017:3).

For some theorists, hate discourse and hate crime are interrelated, targeting one or more innate (identity) characteristics of their victims. Thus, these phenomena seek to send messages of hatred and intolerance against such inherent qualities, causing damage on three basic levels: the micro-level (to the victim), the mid-level (to the victim’s group) and the macro-level (to society, e.g. through the degradation of social cohesion) (Alkiviadou 2018:219).

As we have already discussed, hate discourse as a concept appears as quite omnibus and open to different interpretations. By this logic, ‘narrower’ concepts such as “dangerous speech” and “fear speech” seem, to some theorists, more appropriate, focusing on the ability of discourse to cause harm and lead to violence (Gagliardone et al. 2015:12).

Susan Benesch (2012) has proposed a framework that offers the ***ability to recognize dangerous speech***, based on:

1. the character and popularity of the speaker,
2. the emotional state of the audience,
3. the content of the verbal act as a call to action,
4. the historical and social context in which the verbal act takes place, and
5. the means used to spread the dangerous speech, including language (e.g., if a local dialect is used a message may not have the same meaning for all those who speak the official language).

In addition, the concept of “fear speech” (Buyse 2014) signals the dissemination of messages aiming to create feelings of fear or threat to a specific social group, legitimizing the perpetration of violent acts as a seemingly necessary defensive strategy to ensure the security or integrity of the “threatened” group (Gagliardone et al. 2015:12).

From what has been mentioned so far regarding hate discourse, it should be understood that it constitutes a way of creating value differentiation of people through -any- means of communication (speech, writing, images, etc.) (Sponholz and Christofoletti 2019:68). For this reason, it can be addressed as a (public) communication issue (Sponholz and Christofoletti 2019:69) and not just a linguistic matter, since communication means exchanging, understanding, interacting, while language is merely the system of symbols used for communication ( Sponholz and Christofoletti 2019:68).

Hate discourse has spread in today's online world, without easily traceable sources and causes, and without crystallized methods for achieving accountability and resolution.

At the same time, alongside the development of new media -mainly the Internet and its individual communication applications-, emerged **the phenomenon of online hate discourse** (Antigone 2019:2). Hate discourse has spread in today's online world, without easily traceable sources and causes, and without crystallized methods for achieving accountability and resolution. The comments of mainstream news websites, blogs, and social media have become places of expression and dissemination of hate discourse, rather than a space for dialogue and exchange of arguments (Mollen 2018:1).

In July 2017, the Pew Research Center reported that four out of ten U.S. adults (41%) had experienced **cyberbullying** and many more had witnessed a similar incident. From defamation and public humiliation, to physical threats and harassment, much of this activity is directed against women, trans people, and members of racial or ethnic minorities. The same research shows that one in four black people has been the target of harassment and that women are twice as likely as men to be ‘harassed’ online. In addition, religious and political beliefs are often causes of expressing aggression (Reyman and Sparby 2020:1).

Regarding online hate discourse, Hawdon, Oksanen and Rasanen (2017) argue that its core -in direct relation to the general definitions of hate discourse- expresses (within the Internet framework) the hatred of a collective and is, thus, aimed at identity characteristics of a collective (e.g. race, ethnicity, gender, religion, sexual orientation) (Sobieraj 2019). It originates mainly, but not exclusively, from white men, who are often (but not always) associated with far-right, racist groups (Porter 2020:xx; Jakubowicz et al. 2017:211). During the interviews conducted for the purposes of the present program, several of the interviewees ‘identified’ men, regardless of age, who in many cases are also associated with far-right organizations, such as -in the case of Greece- the Golden Dawn party, as the main perpetrators of hate discourse.

Hawdon, Oksanen and Rasanen (2017) also argue that the direction of hatred towards a collective and its inherent characteristics and not towards an individual differentiates hate discourse from ***cyberbullying***. Therefore, hate discourse includes slanders, religious and political statements of extremists, and displays of hate symbols (e.g. swastikas, burning of crosses) with the goal of denigrating/discrediting certain social groups (Miskolci, Kovacova, Rigova 2018:2).

According to some theorists, one should distinguish between ***annoying but relatively harmless comments (trolling)***, and more serious and systematic forms of hate expression, such as gang bullying, coordinated and persistent attacks, or organized attempts to spread false news. A merely playful statement differs from the constant individual attacks taking place over a specific period of time (bullying), which -in turn- differs from an organized collective attack perpetrated by an entire group of racists or misogynists (Porter 2020:xix). By this logic, hate discourse and cyber-trolling -the attempt to discredit the effective exchange of arguments in the context of an online dialogue-, although theoretically unidentical, coincide in that both can be considered manifestations of rejection of the equal dialogue on which many online communities rely (Mollen 2018:60).

Nevertheless, ***the boundaries between trolling, bullying, and hate discourse are rather blurred***, as Internet trolling can manifest itself with varying intensity each time, so that at least some of its manifestations fall into the logic of hate discourse, while bullying (personal attack) may well hide feelings of hatred against a specific social group (Sobieraj 2019).

Especially on social media, hate discourse can spread very easily and includes posts, comments, or memes (images or videos), posted publicly or in closed groups on different platforms, such as Facebook or Twitter, or in private messages sent to emails of targeted individuals or groups (Miskolci, Kovacova, Rigova 2018:3). The ‘damage’ that hate messages can cause becomes greater as long as they remain posted on the Internet and accessible to users (Gagliardone et al. 2015:13).

The online emergence of hate discourse also means that hate messages that may not have found their way into the public sphere and support in the past, can now be placed in platforms that are visible to a large audience.

Even when web hate content is removed (from a platform), however, it can easily re-emerge elsewhere, possibly even on the same platform under a different name, or on different websites. If a site gets shut down it can be reopened quickly using a less stringent server or in a country with looser hate discourse laws. The online emergence of hate discourse also means that hate messages that may not have found their way into the public sphere and support in the past, can now be placed in platforms that are visible to a large audience (8/13). The ‘durability’ of online hate discourse material is unique due to the low cost of production and distribution and the possibility of immediate ‘revival’, which ensures the seamless presence of hate discourse in different parts of the Internet public sphere (Gagliardone et al. 2015:14): “*Deeply rooted stereotypical views of reality have been transferred to the digital environment of social media. One must keep in mind that comments on the Internet are not easily deleted*” (CJ).

All of the abovementioned approaches -with all their variations and complementary aspects- converge on the description of a phenomenon which occurs often, both online and offline, and can ‘target’ different individuals and social groups, having a ‘preference’ for individuals and social groups that deviate from the prevailing ideals of a society, as they are ‘registered’ and widely accepted in its culture: *“(Online) hate discourse aims to degrade minority groups through speech (e.g. people with disabilities, people with a homosexual sexual orientation, and people with 'divergent' political beliefs)”* (SJ). That last parameter reveals the profound socio-political character of hate discourse and the fact that it is formed within and because of specific social contexts. At the same time, it is a form of discourse that can be manifested with varied ‘intensity’, expressing hatred towards specific individuals and/or groups, both directly (e.g., direct threats against bodily integrity) and indirectly (e.g., disparaging comments, attempts to create scapegoats), in order to cause irreparable mental and/or physical damage to the subjects it targets.

In this regard, there seem to be two different categories of hate discourse. On the one hand, there is what might be called ***hard hate speech***, which includes features that are prohibited by law (e.g. direct threats to life), and on the other hand, there is ***soft hate speech***, which does not incur criminal penalties but raises serious concerns about growing intolerance and discrimination (e.g. derogatory comments about certain social groups - immigrants are
‘dirty’). As we will see later in this text, *the line between hard and soft hate speech* (especially in relation to protected identity characteristics) *varies from country to country*. In addition, different democracies incorporate completely different approaches to regulating and combating hate discourse. Thus, while the US prioritizes the protection of freedom of expression and opinion, many EU Member-States implement legal measures to regulate and combat hate rhetoric (Assimakopoulos, Baiden and Millar 2017:4-5).

Based on the aforementioned approaches, the social constructionism that proposes a socio-ideological dimension of discourse and the corresponding critical analysis that focuses on socially constructed inequalities, ***we could define hate discourse as any -mainly publicly expressed- intentional form of communication (either online or offline) which contains, either directly or indirectly, hatred/disrespect/insult towards specific individuals and/or -by extension- specific social groups, attacking their identity characteristics on the basis of dominant social constructions and values*** ​​(creating, strengthening, and perpetuating social prejudices), in order to devalue these individuals or groups in the perception of society as a whole and, consequently, to marginalize or even eliminate them, with whatever consequences this may have on an individual and social level.

At this point it should be emphasized that due to the fact that hate discourse encompasses a certain ideology, therefore a specific way of constructing and interpreting the social reality, it is understood/identified as such based primarily on ideological criteria. In other words, usually the person who expresses hate discourse based on their own ideology does not realize the extent to which they are, in fact, participating in hate rhetoric (for example, someone who consistently inveighs against immigrants justifies their views by convincing themselves that immigrants pose a -more or less significant- ‘threat’ to his way of life). Therefore, hate discourse will be identified either by an organization/body for the protection of immigrants' rights, or by a person/friend (ideologically) close to the immigrants. In this sense, hate discourse is the (ideological) narrative of a social reality that its exponent evaluates negatively based on personal ideological criteria.

Furthermore, what may be generally deemed acceptable in one place and time may not be equally acceptable in a different social context. For example, the way a society treats the elderly, or women's rights, or defines/condemns far-right ideas is completely different in each country (Heinze 2016). That being said, we need to emphasize the importance of social and political factors that determine what is acceptable or not at any given time in a particular social context (La Macchia and Louis 2016). Of course, no one can ignore the impact that other cultures or universal ideas may have nowadays, as they are being globally spread through websites, blogs, and social media, in producing and facilitating the dissemination of trends and ideas in different countries (Royzman, McCauley and Rozin 2005; Cochliou 2020).

Consequently, the socio-ideological dimension of hate discourse contributes, in its own way, to the difficulty of identifying and dealing with it on a practical level. Due to the fact that ideologically charged expressions are inherent in the symbols system through which communication takes place on a daily basis -i.e. language-, they have become part of Construction, the ‘social DNA’ of a society. For this reason, it is difficult to label them as -direct or indirect- references to stereotypes, which, depending on the ideological position of the subjects (individuals) involved, can turn into prejudices and expressions of hate discourse and/or violent behavior. One such example is the expression often used by mainstream media when referring to current prices of basic consumer goods as “the housewife's basket”. This particular expression, although it is a fixed expression used -at least seemingly- in a neutral way, perpetuates the stereotype of the woman/wife as solely responsible for ‘housekeeping’ a home. Therefore, it may, in view of certain ideological criteria, perpetuate the stereotypical perception that ‘housework’ is the exclusive obligation of women, hence any deviation from this logic can lead to psychological and/or physical violence against women. On a similar context, we also find colloquial expressions, such as the “lame Mary” in Greek *(“κουτσή Μαρία”*, which roughly translates to *“riffraff”*), which can normalize a logic of discrimination against people with disabilities.

Words can hurt just as much as attacks against our physical integrity. Racist, sexist, homophobic, and other related adjectives become “weapons that terrorize, hurt, and humiliate” (Heinze 2016:2). Hate discourse can lead to harmful effects on the social groups or individuals it targets. Researchers in various fields (e.g. medicine, psychology, social sciences, and political sciences) have focused on examining the negative effects of hate discourse. Research has shown that hate discourse negatively affects the (mental) health of victims, threatens democracy by silencing minorities and suppressing their political participation in democratic processes, and therefore strengthens and perpetuates policies of exclusion, and can also incite violence (Miskolci, Kovacova, Rigova 2018:3): “*The internet greatly shapes public opinion. A person who may not have a formed and complete personal opinion on an issue can enter a social networking site, read 15 individual opinions, and consequently form their own. So, if racist views are published without any control, I believe that, in the end, perceptions with racist content are formed and hate rhetoric is expressed. Therefore the consequences are serious”* (YJ)*.*

The extent to which individual hate messages can cause harm depends on how the victims perceive and react to them, which in turn is influenced by the victims' personalities and (social) aspects of their existence, as well as from other factors that may affect the psychology and behavior of the recipients (Strossen 2018:22). In any case, however, there is always the serious possibility that hate discourse will cause significant, even irreparable harm to its victims: “*The consequences for the victims are huge. Studies have shown that the harm suffered by victims of hate discourse is comparable to the harm caused by physical violence”* (CJ).

Hate discourse is not a phenomenon that has only recently appeared or will soon disappear, regardless of the volume of counter-narratives and the tools that will be used to combat it.

In conclusion, hate discourse is at the center of social organization and involves, while simultaneously ‘transcends’ identity characteristics such as class, gender, nationality, religion, sexual orientation, etc. (Antigone 2019:28). Perhaps the most important issue regarding hate discourse is the fact that fundamental (for many, but not for all) human rights and identity characteristics have been ‘ideologized’, so they may become the subject of (negative) evaluative judgement and, therefore, hate discourse. Perhaps, then, the only way to deal effectively with hate discourse is to de-ideologize these parameters and integrate them into a framework of crystallized, fundamental, and unquestionable/commonly accepted social values. The act of labelling the negative -against identity features- discourse as hate discourse could be the beginning.

# 2.3 Summary

In this section we have presented the different approaches around hate discourse and have come up with a definition that undoubtedly emphasizes the social character of hate discourse.

Stereotypes and prejudices are key concepts in examining hate discourse. Stereotypes are the mental processes of categorizing reality based on certain information/characteristics, while prejudices are the emotional projections of (negative) value judgments on these entrenched mental categorizations. Deeply-rooted prejudices are spread through speech in the form of narratives. A narrative is a logical, or at least seemingly logical, and internally coherent interpretation of connected events or information that makes sense to the reader/listener. When a narrative is presented as the only correct one, rejecting alternative narratives or, in extreme cases, inciting violence against those who question it, in other words taking the form of hate discourse, the fundamental rights of a pluralistic society (e.g. the right to freedom of thought, freedom of expression, freedom of religion and belief, etc.) are at stake.

We, then, referred to a multitude of definitions, which we categorized according to the aspect of hate discourse they emphasize the most. Therefore, we referred to definitions formulated by political institutions (e.g. European Commission) which describe the manifestation characteristics of hate speech, we presented definitions that emphasize the purposes of hate discourse (e.g. propaganda) and, finally, settled -through approaches that emphasize the social dimension of the phenomenon- on a definition/description of hate discourse which focuses on its causes, ways of manifestation, and possible consequences: *we could define hate discourse as any -mainly publicly expressed- intentional form of communication (either online or offline) that involves, either directly or indirectly, hatred/disrespect/insult to specific individuals and/or -by extension- to specific social groups, attacking their identity characteristics on the basis of dominant social constructions and values ​​(creating, reinforcing, and perpetuating social prejudices), in order to devalue, in the eyes of society as a whole, the perception of these individuals or groups, and, consequently, to marginalize or even eliminate them, with whatever consequences this may entail at both the individual and social levels.*

We also made reference to the dissemination of hate discourse on the Internet, as hate discourse has spread in today's online world without easily traceable sources and causes and without crystallized methods for achieving accountability and resolution. The manifestation of online hate discourse constitutes a ‘gap’ between the new communication features brought by the interactive character and the digital structure of the Internet and the pre-existing social prejudices which reflect and reproduce power relations, and therefore inequality, at the social level.

# Strategies for combating hate discourse

# 3.1 The legal and ‘self-regulatory’ treatment of hate discourse

“Hate speech needs to be understood […] as a social phenomenon within the context of unequal social and power relations, in which stereotypes, emotions, and narratives play an important role” (Del Felice et al. 2017:29). Online hate discourse, as we have already discussed, is a phenomenon presenting an ever-increasing trend with evolving forms, perpetrators, goals, and strategies (Gagliardone et al. 2015). Regarding the legal regulation of hate discourse, one can perceive different points of view in different sub-sections of this discussion.

In this section we will first present the discussion surrounding the legal and self-regulatory logic of dealing with (online) hate discourse. The main argument is that hate discourse efforts can be effective in many cases, but in reality they only address the symptoms and not the causes of hate discourse. As Andre Oboler, CEO of the Online Hate Prevention Institute, notes, “The longer the (hate) content stays available (on the Internet), the more damage it can inflict on the victims and empower the perpetrators. If you remove the content at an early stage you can limit the exposure. This is just like cleaning litter, it doesn’t stop people from littering, but if you do not take care of the problem it just piles up and further exacerbates” (Gagliardone et al. 2015:13).

First of all, *there is a fundamental disagreement regarding the need for some kind of legal regulation on this issue*. A typical example is the difference of legal treatment between the US and various other ‘Western’ countries (European and non-European), with the former not including a specific legal text that defines and restricts hate discourse (Sobieraj 2019) and the latter having incorporated into their legal systems provisions that criminalize various forms of hate discourse (Waldron 2012; Gagliardone et al. 2015:11). The US logic is based on avoiding imposing any restriction on the freedom of speech as established by the First Amendment (Strossen 2018). Therefore, the spirit of non-criminalization of hate discourse is based on the argument that it should primarily be considered a form of discourse and secondarily an expression of hatred, although US lawyers have expressed the view that any prohibition/criminalization of hate discourse is not, in fact, in opposition to the Constitutional provision of the First Amendment (Sobieraj 2019), precisely because it expresses hatred for the identity characteristics of the subjects it targets, thus degrading their freedom of expression that is at the core of the First Amendment. This same logic can also be found within the legal system framework of several other countries (Canada, France, Germany, etc.), which have enacted specific legal provisions against hate discourse. However, due to the strongly cultural/ideological nature of hate discourse, there are both similarities and differences in the way hate discourse is approached in the legal texts of the countries that favor its legal treatment (Assimakopoulos, Baiden and Millar 2017:10· Jakubowicz et al. 2017:185· Waldron 2012:8-9).

Similarly, fundamental disagreements can also be found in the case of the legal treatment of hate discourse regarding the effectiveness of the methods followed. ***Advocates of the legislative/criminal treatment of hate discourse argue that the state has a responsibility to combat hate propaganda***. Proponents of criminalization are concerned about the possible transmission of hatred, the psychological damage that any minority group targeted by hate discourse may suffer, and the threat to the stability and tranquility of society in the event that hate propaganda succeeds in delivering its messages to the majority (Cohen-Almagor 2012:44). They further support this view by arguing that “in a perfect world, we would respond to hatred with education, not laws. However, our world is not perfect and history has shown that hate speech can lead to horrific crimes” (Cohen-Almagor 2012:56): “*Laws against hate discourse must be enforced. Freedom of speech has to do with protecting the speaker from the oppression of power and not with hate discourse. Therefore, the application of specific laws against hate discourse does not degrade the freedom of expression*” (FGJ).

More specifically, in the context of online communication, the fight against hate discourse can develop on three levels: the level of (daily) content regulation (moderation level), the oversight level, and the regulatory level. At the ***moderation level***, online platforms, or individuals (e.g. volunteer coordinators), or organizations (e.g. subcontractors) participate in the ‘regulation’ of hate discourse, i.e. the comparison of the content in question with “community standards” or “content policies” of online platforms, in order to decide whether the platform will allow, remove, add warning tags, or take any other action in relation to the content. The ‘final’ content regulation decisions may be made by humans, machine learning tools, algorithms, or a combination of these (Brown 2020:31). This practice is a kind of “ex-post self-regulation” solution, implemented after the publication of the message in question (Brown 2017:15).

At the ***oversight level***, the rules, procedures, and acts regulating hate discourse are subject to external or internal control. What matters at this level of control is that there is an evaluation of the content regulation process. This evaluation can be carried out by governmental or non-governmental organizations, by the Internet body itself, by volunteer users, by independent supervisory authorities, etc. (Brown 2020:32).

At the ***regulatory level***, national governments or international governmental bodies are involved in online governance through the enactment of laws and codes of practice, which online platforms must follow in order not to be penalized, following inspection by the police and/or prosecutorial authorities (Brown 2020:32).

The application of such a multi-level and pluralistic logic of online content control, offers -according to the devotees of the legal treatment of the phenomenon- multiple possibilities, both for the benefit of the victims of hate discourse, and for the benefit of the public sphere and society in general (Brown 2020:37, 44)

Nevertheless, cooperation between all these institutional factors (online corporate governments, non-governmental organizations, or other civil society actors) to tackle online hate discourse is not always harmonious. This is because these actors usually have different motives and purposes behind the design and implementation of tools and practices to reduce hate discourse, and these differences are often reflected in the different definitions of the phenomenon (Brown 2020:48) and the disagreements/conflicts between the involved stakeholders, both governmental and non-governmental (Mollen 2018:236).

Much of the burden and responsibility of identifying and tackling online hate discourse has been focused on the governments. ***Attempts to limit hate discourse, however, are not limited to the public, institutional level***. As we have already mentioned, Internet communication service providers, such as companies-owners of social media platforms (Facebook, Twitter, etc.), have tried to curb hate discourse, playing an increasingly important role in both enabling and tackling hate rhetoric. Many of the online providers, especially search engines and social networking platforms, operate supranationally and regulate user interactions on a daily basis based on their own operating agreements, terms of use, and definitions of hate discourse (Brown 2017:21).

***Internet companies have developed different definitions of hate discourse and guidelines for regulating it.*** Some companies do not use the term “hate speech”, but have a list of terms related to it (Gagliardone et al. 2015:29). For example, Facebook prohibits hate speech, which is defined as: direct and serious attacks on any protected group of people based on race, nationality, ethnic origin, religion, gender, sexual orientation, disability, or illness. Platform officials have stated that “we are working hard to eliminate hate speech quickly, but there are cases of offensive content, including unpleasant humor, that are not considered hate speech based on our definition” (Jakubowicz et al. 2017:309). Despite investing in industry-leading systems to detect and eliminate hate discourse, Facebook has acknowledged that: “It has become clear that our systems for detecting and removing hate speech content have failed to function as effectively as we would like ... [and] the content is not removed as quickly as we would like” (Jakubowicz et al. 2017:310).

In addition, online companies rely heavily on user reports of content deemed inappropriate (Gagliardone et al. 2015:53) and when such a case is identified, it is assessed -at least initially- according to their own terms of operation and not based on a specific legislative framework. Additionally, the conditions under which online communication intermediaries operate, in terms of how these are related to national laws and international regulations, pressure groups and individual users, are constantly changing (Gagliardone et al. 2015:55). Although they have been repeatedly asked to make their content regulation process more transparent, social media platforms have so far avoided proposing stricter rules and procedures in order to identify content that should be removed (Gagliardone et al. 2015:53). ***Perhaps the greatest challenge is not so much the way in which online companies define hate discourse, but how (and, even, if) they implement hate discourse restraint policies***, as many companies claim that they do not create or control online content and should, therefore, have limited liability (Gagliardone et al. 2015:30).

***The issue becomes more complex when considering global social networking platforms*** (e.g. Facebook, Twitter, Snapchat). Given the sheer volume of data they handle, social media platforms rely heavily on user reports regarding content they deem inappropriate, offensive, or dangerous. The platforms then decide, mainly according to their own terms of service, whether this content should be removed, or whether other measures should be taken to restrict access to it or limit the ability of its creators to continue using the services of this platform (Gagliardone et al. 2015:31).

Critics of the legal regulation of hate discourse argue that the laws, although well-intentioned, do little to address the real dangers of hate discourse, because they apply a private and personalized process of resolving an essentially public/social issue, which means that there is little room for the immediate confrontation of the wider social causes and consequences of hate discourse. At the same time, any -personalized- resolution of the problem is left to individuals who are willing to file a formal complaint: “*The victims feel helpless (to react), as there is no specific legislation. The police are ineffective unless they receive a report for a direct threat to life.”* (CJ). Additionally, the laws in several cases propose a rather restrictive definition of hate discourse, thus resulting in several cases of hate discourse falling just outside the legal framework (Heinze 2016:4; Gelber 2002:1; Gagliardone et al. 2015:16; Brown 2020:48-49). Hate discourse legislation can also ‘force’ its actors to ‘disappear’, depriving society of the opportunity to record hate speech perpetrators, while some scholars argue that those expressing hate rhetoric also miss the ‘opportunity’ to defuse some of their hatred verbally, before transforming their beliefs into acts of physical violence (Brown 2017:17; Cohen-Almagor 2012:44).

As stated by the UN Human Rights Council Special Rapporteur on minority issues, “It is true that hate crimes rarely occur without prior stigmatization of targeted groups and incitement to hate incidents”. At the same time, “only the most horrific forms of hate rhetoric, that is, those that incite discrimination, hostility, and violence, are generally considered illegal”. The Rapporteur goes on to note the importance of distinguishing between three types of hate discourse (Gagliardone et al. 2015:15-16):

1. a form of expression which constitutes an offense under international law and may be prosecuted,
2. a form of expression which is not criminally punishable but which may justify certain restrictive measures,
3. a form of expression that does not entail criminal or civil penalties, but still raises concerns about tolerance and respect for others.

 He particularly emphasizes that, while there is the possibility for legal action regarding type (a) hate discourse, legal measures cannot be considered as an adequate response to the full range of hate discourse that can create the background for the perpetration of hate crimes.

Apart from the ‘purely’ legal issues, there is also the entanglement of the law with the government. From a sociological or anthropological point of view, we know that the law imposes the morals of the legislative elite that can control the content of the law and, consequently, exploit it at will (Strossen 2018:6). For example, the hate speech law during South Africa's Apartheid was used to criminalize criticism of white supremacy, highlighting the possible political exploitation of any legal measures against hate discourse (Gagliardone et al. 2015:15). The government's ability to ‘neutralize’ ideas that are unpleasant and disturbing not only undermines freedom and democracy, but also undermines the objectives of equality and justice that hate discourse laws are supposed to pursue. Such laws apply to the suppression of unpopular speakers and ideas and, too often, even stifle the rhetoric of vulnerable, marginalized minority groups which they are supposed to protect (Strossen 2018:15).

Based on the above problematic points of the legal regulation of hate discourse, some theorists argue that precisely because the crime of hate discourse is an extremely problematic offense, and because censorship burdens democracy, it is better to respond to the threat of hate discourse encouraging more dialogue, rather than through criminal sanctions (Cohen-Almagor 2012:43; Gelber 2002:1). Besides, we must not overlook the fact that any (legal-regulatory) intervention on public speech expression constitutes merely an ex-post intervention (treatment), which may be able to ‘regulate’ the problem, but does not eliminate it. Any radical treatment of this phenomenon should be aimed at preventing its occurrence, identifying its socio-cultural causes and ‘neutralizing’ it through the gradual addition of the need to respect individual and social human rights, and -in this case- those linked to people's identity characteristics, the social DNA (Construction).

In 2015, the **European Commission Against Racism and Intolerance** (ECRI) published a report criticizing European hate speech Laws as ineffective and potentially counterproductive; the report called on European countries to prioritize non-censorship alternatives, including responding (to hate speech), concluding that these measures are “much more likely to prove more effective in the eventual elimination of hate speech and its potentially harmful consequences, than censorship”. On the same year and in the same spirit, UNESCO published a report on addressing hate discourse, which advocated the logic of “response” as opposed to “oppression” in tackling hate speech (Strossen 2018:24). In 2017, the German-based European Center for Press and Media Freedom (ECPMF) issued a statement against the proposed German legislation requiring social media to block or repeal hate discourse, explaining that the “fight against illegal incitement to violence, hatred, and discrimination is indeed crucial, but censorship of expression has never been effective. It is likely that our societies can be helped substantially through creating more dialogue” (Strossen 2018:24).

The progress we have already made by encouraging more discourse, not less, should encourage us to continue on this path. We should all exercise our most crucial right to promote this vital purpose: the right not to remain silent (Strossen 2018:186): “*Not speaking can also be wrong, because this allows those who express hate discourse to prevail, while the wisest and calmer voices that could face these issues are diminishing*” (YJ).

The description of what constitutes hate discourse and the regulations that should be applied to combat it are ‘problematic’ points, as the legislation differs from country to country. The definition and classification of hate discourse depends on the content and tone of expression, the intent of the speaker, the communication context, the purpose of the expression, and its impact on the targeted person or group. The 2015 UNESCO Report on Online Hate Speech recognizes these challenges and also points to the challenge of establishing concrete definitions in practice when boundaries are often blurred.

Despite any problems, defining hate discourse as a violation of fundamental human rights and combating it through legal means is imperative. **Hate discourse may be identified and sanctioned under applicable law**, yet it should also be understood as a worrying social phenomenon that erodes the foundations of democratic societies, such as respect for diversity, and we should focus on treating its root causes, otherwise only the ‘symptoms’ can be alleviated (Del Felice et al. 2017:33).

## 3.1.1 The international institutional framework

As Cochliou (2020) points out, freedom of expression has been recognized as one of the most fundamental human rights and is, therefore, enshrined in every relevant international Treaty. The absolute right to freedom of expression has certain restrictions, which are protected by the Treaties. Despite the “provisions” of the Treaties regarding freedom of expression, as far as online hate discourse is concerned, most Treaties do not include direct references. Nevertheless, key international human rights protection instruments, such as the Universal Declaration of Human Rights (UDHR 1948) and the International Covenant on Civil and Political Rights (ICCPR 1966) address issues of social inequality and racial or other discrimination, constituting basic human rights maps and a ‘compass’ for other international or national legal texts.

The Universal Declaration of Human Rights constitutes the basis for all international human rights law, although it is not a binding legal text. Article 2 states that:

“Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty”.

In addition, Article 7 provides that: “All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination”. Freedom of opinion is further protected by Article 19: “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers”.

On December 21, 1965, the UN General Assembly adopted the Convention on the Elimination of All Forms of Racial Discrimination (CERD). First, Article 1, paragraph 1 sets out the definition of racist discrimination as:

“Any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life”.

Additionally, Article 4 sets out the principles of racism and freedom of expression:

“States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention”.

The Convention then sets forth explicitly that States must punish any dissemination of ideas based on racial criteria, as well as prohibit any activity and funding of the relevant organizations.

Although there is a strong background in international law regarding human rights treaties and conventions, Member-States, including some of the most developed countries, do not protect their citizens from racial discrimination, intolerance, hate speech, and so on. Therefore, the rhetorical question that naturally arises is whether States are doing enough to protect human rights against social inequality and human rights violations.

# 3.1.2 European institutional framework

In addition to the UN Treaties, the European Union (EU) has developed its own anti-discrimination case law, through the European Convention on Human Rights (ECHR 2010) and through its legislation. The EU also promotes the principle of equality and non-discrimination as a fundamental element of several Council of Europe (EU) Treaties, while protection against all forms of discrimination constitutes a key objective of the Council when developing new rules and regulations. The ECHR was one of the first human rights treaties, inspired by the Universal Declaration of Human Rights, and calls on EU Member-States to recognize and respect fundamental human rights. By that logic, Article 14 prohibits all forms of discrimination:

“The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status”.

In addition, Protocol 12, which was incorporated in 2000 into the same Convention, extends the prohibition of discrimination to the benefit of any legal right, even in cases where that legal right is not protected by the Treaty, assuming that it is provided for based on national law. This guarantees equal rights under both the Convention and national law.

Furthermore, the European Social Charter, adopted in 1996, protects the right of all workers to equal opportunities and equal treatment with regard to work, prohibiting discrimination on the grounds of sex. Additionally, the Framework Convention of the COE for the protection of national minorities, issued in 1994 and put into force in 1998, consists of anti-discrimination provisions, as well as the COE Convention on Action against Trafficking in Human Beings and the COE Convention on Access to Official Documents. Similarly, the Protocol on Xenophobia and Racism (2015) in the COE Convention on cybercrime contains terms related to the distribution of racist material through social media and the Internet and is, therefore, an important tool in the fight against online hate discourse.

Respectively, the Committee of Ministers of the COE adopted, on October 30th 1997, Recommendation No. R(97)20 to Member-States on “Hate Speech”. As already mentioned, the Recommendation includes a definition of hate speech and condemns all forms of expression that incite racism, xenophobia, antisemitism, and intolerance, and highlights the need for Member-State Governments to establish guidelines on how to address these forms of expression. It is worth noting that the recommendation states that such modes of expression can cause more harm when circulating through the media. In addition, the Committee of Ministers adopted Recommendation No. R(97)21 to Member-States on the media and the promotion of a culture of tolerance (1997). The Recommendation states that the media could make a significant contribution to combating xenophobia by promoting the idea of ​​respect and multicultural understanding.

Nonetheless, the European Court of Human Rights has received numerous complaints of breach of Article 10 of the ECHR and has concluded that:

“[T]olerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance …, provided that any ‘formalities’, ‘conditions’, ‘restrictions’ or ‘penalties’ imposed are proportionate to the legitimate aim pursued.” (Erbakan v. Turkey judgment of 6 July 2006, § 56).

The European Union (EU) has adopted a large number of legal provisions against discrimination to be implemented by the Member-States. The EU Charter of Fundamental Rights is based on the ECHR and includes provisions on freedom of expression (Article 11) and non-discrimination (Article 21). The 1999 Amsterdam Treaty gave the EU Parliament the power to adopt methods of combating discrimination, such as the implementation of binding legislation on discrimination based on sex, race, ethnic origin, religion or belief, disability, age, or sexual orientation. To this end, the EU adopted four directives based on Articles 153 and 157 (formerly Articles 137 and 141) of the Treaty establishing the European Community (equal rights for men and women in the workplace and equal pay for men and women)[[2]](#footnote-2).

Based on the above, EU legislation and its Institutions are quite progressive in relation to discrimination, providing protection to specific groups with the goal to reinforce equal access to the labor market, as well as to prohibit discrimination in access to goods and services. All EU Member-States have incorporated the relevant directives into national law, while in most countries there are laws that protect citizens from social discrimination in a broader context than that provided by the EU directives. However, there are several important gaps in the application of these laws.

As is evident from the above, in the EU, there is no specific law that deals with online hate discourse. The Charter of Fundamental Rights of the European Union does not explicitly refer to hate speech in any of its provisions. However, it protects both freedom of expression (Article 11) and the right to non-discrimination (Article 21), which must be seen as the basis for protection against hate discourse. In addition, the fight against racism is a constant objective for the European Union. To this end, the EU implemented a strategy against any form of racism and discrimination, clearly reflected in the adoption of the Framework Decision on combating certain forms and expressions of racism and xenophobia through criminal law[[3]](#footnote-3). Under this Decision, Member-States shall take all appropriate measures to penalize offenses which constitute hate speech, as described in the Decision text, i.e. public incitement to violence or hatred directed against a group of persons or a member of a group of persons, which is defined on the basis of race, color, origin, religion or belief, or national or ethnic origin.

In an effort to regulate the digital sphere, which is characterized by inherent complexity with regard to any legal regulation effort, the EU has turned to more lenient legal measures. In May 2016, the European Commission agreed with Facebook, Microsoft, Twitter, and YouTube on a “code of conduct to tackle illegal online hate speech”[[4]](#footnote-4). This code has laid the groundwork for a set of measures to be implemented by technology companies to tackle hate discourse on their platforms. In addition, companies are required to introduce a set of rules defined as “Rules or Community Guidelines”, which prohibit individual users from using their platforms to incite violence or spread hatred. In practice, the templates described in the code do not, in fact, bind online platforms. Instead of participating in the Code, Internet service providers agree to self-regulate activity on their respective platforms accordingly on a purely voluntary basis. Over the years, Instagram, Google+, Snapchat, Dailymotion, and Jeuxvideo.com have joined the Code. This Code is evaluated for its effectiveness on an annual basis[[5]](#footnote-5).

It is, therefore, clear that the EU is trying to establish harmonized standards for combating online hate discourse and hate crime in general, but the involvement of national and community legislation in relation to the global nature of the Internet results in a lack of specific, concrete legislation on the subject and the consequent ‘problematic’ regulation of the field.

## 3.1.3 Online hate discourse and the legal framework in Greece and in Cyprus

In both Greece and Cyprus, online hate discourse is a growing problem for which there is no exact method of tracking it down. In both countries this can be attributed to three general factors:

1. The biased social attitudes towards specific groups of people with reference characteristics of race, color, religion,
2. The evolution of online media and the lack of user awareness about the fact that they may be involved in hate discourse,
3. As well as the lack of appropriate tools for recognizing and tackling hate discourse.

***Both Greece and Cyprus are above the European average in terms of xenophobic behavior.*** In Greece, for a number of years and in conjunction within the economic crisis and the prevailing refugee issue, racism and xenophobia have been on the rise (Eurobarometer 419 2014; Racist Violence Recording Network 2018; ECRI 2016). In Cyprus, the majority of hate discourse cases in the public sphere are addressed towards refugees, immigrants, and Turkish Cypriots, while there have been several cases of journalists reproducing hate discourse, which originates from government officials and members of parliament (In-Cyprus 2020).

At the same time, ***stereotypes about sexual orientation and gender identity prevail in both countries*** (Eurobarometer 437 2015), creating fertile ground for hate discourse expression against the LGBT+ community. The recent murder of Zak Kostopoulos, an activist of the LGBT+ community, who was beaten to death in broad daylight while trying to escape a jewelry store, and his subsequent representation by the Greek media as a “drug addict robber” are indicative of social discrimination based on sexual orientation.

In both Greece and Cyprus, the rise of blogging and social media (mainly Facebook) has affected the way people interact with the news and, consequently, the way the news are written and reported. In Greece, a country with high political polarization and a long ‘tradition’ in media intstrumentalization (Hallin & Papathanassopoulos, 2002), it seems that the advent of social media has affected the news environment at various levels. Due to the rapid pace of news production and consumption, journalists often sacrifice credibility and accuracy for commercial purposes.

In addition, ***Greece does not yet have an appropriate legal framework that specifically refers to hate discourse***. However, in September 2014 the Greek parliament ratified law 4285/2014, which is most commonly known as the ***anti-racism law***, under which hate discourse that incites criminal activities can, henceforth, be prosecuted. Human rights activists do not refrain from expressing their concern about this law, however, because of certain “hidden risks”, as the redefinition of the principles of free expression it encompasses may, in fact, restrict the right to free expression itself. Nevertheless, ***the anti-racism law has been deemed innovative in the sense that it has broaden the definition of discrimination, so that manifestations of racism and xenophobia may be considered a criminal offense***. Prior to the anti-racism law, racism and other forms of discrimination were regulated by various legal texts in the national legal framework as well as by international conventions ratified by Greece.

In the case of Cyprus, the legal treatment of certain forms of racism and xenophobia through the penal code began in 2011 with Law 134(I)/2011, which incorporated the Decision 2008/913 of the Council of Europe. Any reference to hatred within this Law concerns “hatred based on race, color, religion, decent, or ethnic origin”. In addition, Article 3, paragraph 1 of the same Law provides that: “Any person who intentionally broadcasts publicly and incites, in any way, violence or hatred against a group of people or a member of a group of people, as defined on the basis of race, color, religion, decent, national or ethnic origin, in such a way as to provoke public disturbance or in a threatening, abusive, or offensive manner, is subject to imprisonment of up to five (5) years, or a fine of up to ten thousand euros (10,000), or both in case of conviction”.

In addition, Cyprus has ratified the Additional Protocol to the Convention on Cybercrime on the criminalization of racist and xenophobic acts committed through information and communication technologies such as websites, blogs, or any other form of social media networks, which stipulates that “Member-States are required to criminalize the dissemination or in any other way distribution of racist and xenophobic material to the public through a computer system” (Law 26(III)/2004).

Article 4 of Law 26 (III)/2004, which ratifies the Protocol, provides that: “If a person intentionally disseminates, or in any other way distributes through computer systems, xenophobic or racist material that incites or promotes prejudice based on racial differences or hatred or violence, that person could be sentenced to 5 years imprisonment, or a fine of up to 35,000 euros, or both”. Online hate speech might be considered included in Article 3, paragraph 1 of Law 134(I)/2011, even if it is not explicitly described, especially if this Article is read in conjunction with Article 4 of Law 26(III)/2004.

Both professional as well as citizen journalists produce and reproduce content that contains hate discourse both deliberately, in order to gain attention and visibility, and/or inadvertently, because they are not able to recognize it as such.

Based on the above information regarding the legal texts that focus on hate discourse both -mainly- offline and -mostly indirectly- online, it is understood that ***there is a critical body of legal texts, which places hate discourse outside the framework of free expression, due to its inherently derogatory nature for individuals and social groups, which violates the principle of equality and equity***. Nevertheless, the conservative social values ​​(Construction) that prevail in both Greece and Cyprus, as well as the social and economic problems that have ‘settled’ in the conservative social structure, imply the ineffective treatment of online hate discourse that is also reinforced -among other things- by the media and the journalists. Both professional as well as citizen journalists produce and reproduce content that contains hate discourse both deliberately, in order to gain attention and visibility, and inadvertently, because they are not able to recognize it as such. Thus, they end up amplifying the voices of hate propagandists. Established journalistic practices, such as specific storytelling and writing styles, as well as clickbait culture, contribute significantly to this phenomenon.

The emergence and spread of hate discourse on the Internet is an evolving phenomenon and collective efforts are required to understand its significance and consequences, as well as to develop effective solutions. The establishment of stricter penalties for online hate discourse constitutes only a first step. ***However, the exclusive focus on restrictive-countering measures ignores the complexity of a phenomenon that is still poorly understood in the context of society and which requires adapted and coordinated responses from a number of different actors in society***. Websites, given their ability to facilitate interaction and provide unprecedented amounts of data that can be analyzed through a variety of new techniques, present a unique ‘magnifying glass’ to examine human behavior. Effective solutions must be based on a better understanding of how different forms of hate discourse appear, interact, and may disappear in this environment (Gagliardone et al. 2015:53). Only in this way will we be able to address this issue preventively -before it even occurs- and not correctively, when at least some ‘damage’ has already been done.

# The importance of civil society and education in tackling hate discourse

# 3.2.1 User communities and civil society actors

International human rights laws must be regarded as a critical framework for the protection and respect of human rights in the fight against (online) discourse expressing hatred, insult, or defamation (Special Rapporteur on the promotion and protection of the freedom of opinion and expression 2019:21). As already discussed, part of countering (online) hate discourse concerns the legal regulation and/or self-regulation of (online) content, either by public national or supranational bodies, and/or by private companies active in matters of web content.

In addition to regulatory and legal action, we also referred to two more parameters of dealing with hate discourse: ***the need to “empower” citizens*** so that they can articulate their opposing stand to (online) hate discourse and ***the need to educate citizens***, beginning from a very young age, and, therefore, the very start of their socialization course, so that certain parameters of the social Construction (the social DNA - dominant culture) will gradually change and hate discourse can be ‘replaced’ by a -as far as possible- generalized respect for fundamental individual and social rights. Both of these parameters, at the core of their existence, include the concept of education at different stages of people's lives and in different ways. It is important to acquire sufficient knowledge about human rights, what exactly they are, how they are protected by national and international treaties and other tools for defending human rights. In the context of hate discourse, it is -first of all- important to know which human rights are being violated and which national and international legal tools should be used to address it (Del Felice et al. 2017:18).

Scholars and activists against (online) hate discourse suggest a variety of strategies and tools to reduce its prevalence on the Internet. These strategies can be categorized into activities including:

1. educational programs,
2. cultural events and community activities,
3. mainstream media initiatives, and
4. social media and online discussion forums.

These strategies include ***educational, training and critical thinking development activities*** that help expose prejudices and falsehoods in children, young people, or journalists, organizing large multimedia campaigns and even participating in Facebook profile shut downs (Miskolci, Kovacova, Rigova 2018:3). Besides, as relevant research has shown, ignoring hate discourse can lead to its dissemination in various forms (against women, minorities, sexual preferences, etc.) (London et al. 2020:66).

Nonprofit organizations, public administrators, the media, politicians, and Internet service providers, as well as any individual social media user can participate in efforts to combat hate discourse. In addition to engaging in hate discourse reporting and identification, users can actively intervene and interact with those who spread hate speech. Interference in the online world represents social control, which means that some Internet users may impose informal rules (as well as penalties) on others in case someone violates the rules of a website. Therefore, any Internet user can impose informal rules and, at the same time, be subject to them (Miskolci, Kovacova, Rigova 2018:3).

The responses and reactions to hate discourse in cyberspace, both by witnesses and by its targets, are especially important. More than half of those having witnessed hate discourse incidents in cyberspace responded to hate discourse. This type of response is higher on the Internet than in the offline world on a global level. In addition, cybercrime targets of racist attacks are more likely to answer back to their perpetrators compared to the respective offline world incidents, where two-thirds of interviewees do not respond when they are confronted with hate discourse. It is also worth noting the differences between the responses of targeted individuals and persons who have witnessed hate discourse incidents. Hate discourse targets in cyberspace are more active than witnesses, with more than two-thirds of targets actively responding to racist content (Jakubowicz et al. 2017: 90-91).

The lessons that can be learned from these studies suggest that there are some basic actions one should take when responding to hate discourse on the Internet:

1. ***Approaching the administrators of the platform*** that is responsible for posting hate content on the Internet, regardless of the chances of success. Most major social networks have now incorporated reporting mechanisms as a result of public pressure and criticism.
2. ***Documenting and reporting cyberspace hate instances*** to relevant authorities and organizations that specialize in researching online hate groups, such as the Online Hate Prevention Institute or specific community-based groups, and
3. ***Contacting organizations (e.g. NGOs)*** that can apply pressure on behalf of the victims and provide support in an effort to alleviate the emotional or psychological impact of online hate and cyberbullying (Jakubowicz et al. 2017:186).

Civil society organizations are, therefore, important factors in combating this problem. They can work along with Internet service providers to remove offensive content, they can pressure the state to effectively apply the law -where it exists- and they can work collaboratively with the industry to act against hate discourse. They can also support victims in coping with the negative impact that hate discourse can have on their lives. They can help society identify hate discourse, understand its impact, and significantly contribute to recovery efforts from its negative effects (Jakubowicz et al. 2017:187).

Furthermore, identifying online hate discourse is vital in tackling its dissemination (Jakubowicz et al. 2017:212). ***There are different ways to create online communities of resistance and solidarity***:

* Grassroots initiatives
* Civil society organizations associated with communities targeted by hate discourse.
* Civil society organizations dealing with racism either online or offline.
* Civil society organizations representing communities or groups of communities, and
* Government agencies initiatives (Jakubowicz et al. 2017:247).

At the same time, there is a ***wide variety of (online) communities*** that can be formed, such as:

* Permanent communities made up of people who are interested in resisting racism and therefore stand in solidarity against it.
* Temporary and special purpose communities that focus on a specific instance, a campaign, or a specific type of content, such as a news story.
* Communities that focus on one specific type of online hate discourse.
* Communities that deal with online hate discourse in general.
* Communities that are committed to combating hate discourse, either online or offline, or
* Communities with a wider focus -on hate discourse-, which occasionally engage in hate discourse resistance activities and acts of solidarity with victims of hate discourse (Jakubowicz et al. 2017:248).

Another ‘key’ to tackling (online) hate discourse as effectively as possible is ***cooperation between different organizations and communities***, which can significantly increase the effectiveness of actions against hate discourse. (Jakubowicz et al. 2017:248).

Civil society organizations dedicated to anti-racist action in general may have the primary objective of building an online community of resistance and solidarity. A community focused on multiple types of hate discourse in cyberspace benefits from its accumulated experience in using different approaches and tools to deal with different types of hate discourse. At the same time, the various communities are empowering individuals to put up more resistance when their own community is attacked, and to turn that support into solidarity for other communities that are also being attacked at a later stage (Jakubowicz et al. 2017:249).

Communities can also be strengthened through participation of the public and other interested actors. Formal or informal collaborations can be established between different organizations/bodies with similar goals. In addition, online communities are not geographically limited and can be extended to reach and include like-minded individuals and groups from around the world. Recognizing the value of building successful online communities of resistance and solidarity and supporting civil society organizations pioneering the fight against hate discourse is vital to combating online hate expression (Jakubowicz et al. 2017:250).

# 3.2.2 The key role of different forms of education

Different researches have highlighted the potentially ***key role that education can play***, mainly in the medium-long-term treatment of (online) hate discourse (Assimakopoulos, Baiden and Millar 2017; Gagliardone et al. 2015; Gelber 2002; Del Felice et al. 2017). This finding also emerged from a research conducted in parallel both in Greece and Cyprus for the purposes of this program. Among the main results of the semi-structured interviews and focus groups conducted in Greece, it emerged that “the only substantive response to the dissemination of hate discourse is education in its various forms (in kindergarten, primary, secondary, and tertiary education, independent seminars). Lifelong learning is proposed as the most appropriate way of addressing [this phenomenon] in view of the legal regulation of the phenomenon” (Cochliou 2020). A form of education that would create a ***‘network’ of cultural-educational activities*** in order to spread the values of equality, respect, and justice, and thus gradually eliminate hate discourse.

The research findings are similar in the case of Cyprus as well, where it was proposed to set up an independent organization that will take on the responsibility of educating different social groups on hate discourse issues, focusing on journalists, given their crucial role in disseminating information to society. Nonetheless, it is also noted that any educational process is a long and difficult endeavor, with no guaranteed success (Cochliou 2020).

Citizen education focuses on preparing individuals to become responsible citizens through the study of their obligations, rights, and freedoms. One of its main objectives is to ***raise awareness*** about the political, social, and cultural rights of individuals and groups, including freedom of speech and the responsibilities and social consequences that result from it. Research shows that ***awareness of the consequences*** and the means by which racism functions can have a limiting effect on people's willingness to reinforce it through repetition and transmission. The basic social skills required to reduce the likelihood of expressing racist discourse include ***empathy and civility in communication*** (Jakubowicz et al. 2017:344). In some cases, the effective argumentation and skills required to articulate personal beliefs and opinions with respect to fellow citizens are included among the desired learning outcomes set out in the relevant educational programs (Gagliardone et al. 2015:46)

The focus of public education on hate discourse is twofold, as it includes the necessary knowledge and skills to identify hate discourse, while also encouraging individuals to deal with, and confront hate messages. One of the current challenges this type of education faces is adapting its goals and strategies to the digital world, providing not only argumentative articulation skills, but also technological knowledge and skills that a citizen may need to deal with online hate discourse (Gagliardone et al. 2015:46). This can be achieved through a ***combination of social and media education***, i.e. through proper education both in basic social parameters and values ​​-as already mentioned- and in the ways in which the media functions as a social institution.

A common denominator of several such initiatives is the ***emphasis on the development of critical thinking***, in line with the respect for individual and social rights of use of social media, to combat online hate discourse.

The expectation is that this ***knowledge surrounding media structure, function, and projected content*** can enhance the ability of individuals to detect and challenge online hate content, understand certain outcomes, recognize their biased opinions and prejudices, and encourage the elaboration and dissemination of arguments to address this phenomenon (Gagliardone et al. 2015:48).

As mentioned above, education on these matters should start at an early age, hence ***we should teach students to navigate the complex world of the Internet and social media*** and be aware of the rhetorical and ethical complexities of these online spaces.

We must also ***teach students to communicate with respect for fundamental individual and social rights***, but also teach them how to protect themselves when attacked.

We must ***cultivate in them the concept of respect***, teach them to respect others but also to demand respect from others: *“Any relevant educational program should be implemented during the first years of education and be part of the formal education system. For example, sociology should be part of primary education in order to teach [students] both the similarities and acceptance of differences between different cultures”* (FGJ).

We need to teach them ways to protect their own (online) public spaces.

We must ***help them understand why the ethics of freedom of expression*** is vital to equal and just social coexistence (Porter 2020:xxi).

There are three ***general educational goals*** that can be set to address online hate discourse, including its detection, analysis, and action against it.

**“Detection”** helps users comprehend the communication context and, in case they transmit it to their contact networks; **“analysis”** supports users in reflecting on what kind of discourse they encountered, in order to understand how hate discourse is structured and what it is intended for, especially considering the normalization of hatred in the daily online experience of users; **“action”** thenoffers the necessary motivation for users to respond to hate discourse in order to deconstruct it, while challenging its normalization through the media content (Jakubowicz et al. 2017:318-319; Gagliardone et al. 2015:50).

These three goals can be seen as a continuum that transforms progressive objectives into specific objects, each focusing on a different aspect of the problem and offering alternative response approaches to online hatred (Gagliardone et al. 2015:50).

**Table 1: Detection, analysis and action against hate discourse (Gagliardone et al. 2015:50).**

|  |  |  |
| --- | --- | --- |
| **Detection** | **Analysis** | **Action** |
| Spreading awareness regarding hate discourse and its consequences | Recognition and evaluation of hate discourse | Response to hate discourse |
| Presentation and dissemination of information | Analysis of causes, underlying conclusions, and prejudices | Composition of messages against hate discourse |
| Communication of the relevant legal framework | Recognition of discriminatory behaviors | Changing the content of hate discourse |
|   | Report and ‘exposure’ of hate discourse | Monitoring and analysis of Media content |

The first educational goal focuses on conveying information about hate discourse, the second on the critical analysis of the phenomenon, while the third encourages individuals to take specific actions. Initiatives taken on the field of information as an educational goal include ***raising awareness*** about online hate discourse, its different forms and possible consequences. They also provide information on relevant national, regional, and international legal frameworks (Gagliardone et al. 2015:51).

The second educational goal focuses on the analysis of online hate discourse. This analysis involves evaluations of the various types of online hate discourse, including racism, sexism, and homophobia, as well as the numerus other forms in which it occurs. An important aspect of the analysis is the ***critical examination of hate discourse*** in order to identify its causes and to understand the underlying prejudices that fuel it. This analytical process allows individuals to effectively detect and report online hate content (Gagliardone et al. 2015:51).

Finally, the third educational goal identified in educational initiatives focuses on ***promoting actions*** that can be taken to combat hate discourse. Proposed actions vary depending on the focus of the project and the organization, however the main focus is on empowering individuals to respond and combat hate content.

Supporting and empowering hate victims to respond is very important, as in many cases victims choose not to speak due to intimidation or the belief that no one will take them seriously. In other cases, victims may indeed respond to hate discourse, but their response might not achieve -for various reasons- the desired result (Gelber 2002:86). Examples of such initiatives are the training sessions for bloggers, journalists, and activists conducted by the “No Hate Speech” movement, the educational materials and courses developed by “MediaSmarts”, and the media surveillance policies proposed by the “In Other Words” project (Gagliardone et al. 2015:51): “*The topics of hate discourse should also be part of the journalism courses at the university journalism departments. Training should be an ongoing process and not an individual program*” (FGJ).

The “No Hate Speech” movement has organized training sessions for bloggers and young activists, allowing them to discuss in a friendly environment some of their experiences with online hate discourse, and share the best practices to combat it. For example, how participants can best implement communication campaigns that portray minority groups in a positive way to prevent discrimination, or how they can campaign to report and limit discriminatory content and activity on the Internet. The training sessions aim to promote a deep understanding of hate rhetoric and to raise awareness about the impact that bloggers and activists can have on tackling hatred. (Gagliardone et al. 2015:50). An educational approach focused on developing the necessary skills to address hate discourse could be designed to provide victims of hate discourse and their supporters with tools, educational and institutional support in responding to hate discourse. This would allow the victims to challenge the ‘silencing’ that hate discourse attacks aim to achieve, by reinforcing freedom of expression, and to refute -with their own discourse- the claims made by exponents of hate discourse (Gelber 2002:89).

On the other hand, the “In Other Words” program has sought to influence and encourage policymakers and civil society actors to monitor and analyze media activity. In addition, this program supports the presentation of reliable information about minorities and vulnerable groups in media representations, encouraging the monitoring of media content by society, in order to avoid the spread of stereotypes, prejudices, and other forms of discrimination (Gagliardone et al. 2015:50). A policy designed to facilitate ‘response’ to hate discourse maximizes freedom of speech and, at the same time, helps to overcome the debilitating, marginalizing, and enervating effects of discrimination resulting from hate speech incidents (Gelber 2002:135).

There are ***two main rhetorical strategies for dealing with hate discourse***. First, ***the user can address hate discourse through presenting reliable information that refutes myths and prejudices***. At this point it is very important to spread the culture of systematic fact-checking of public information, which is an increasingly widespread tool for dealing with misinformation and the general dissemination of unreliable information in the public sphere (Acula et al 2018; Kalsnes 2018; Lim 2018; Mantzarlis 2018).

Second, ***the user can rely on personal experiences to dispel stereotypes and hatred***. For example, experiences with minority social groups may highlight the fact that the ‘inherently negative’ characteristics attributed to such social group also appear in members of the social majority. Nevertheless, engaging in a dispute in order to confront hate discourse has its difficulties, as those who undertake the task of responding to people expressing hatred are consequently more exposed to similar attacks with hate content (Miskolci, Kovacova, Rigova 2018:4).

**Any response to (online) hate discourse can be structured in the form of “counter” and “alternative” narratives.** “Counter” and “alternative” narratives tackle hate discourse by deconstructing the narratives on which it is based. They also propose (alternative) narratives based on human rights and democratic values, such as transparency, respect for difference, freedom, and equality. They can do this by providing alternative and accurate information, using humor, inspiring emotion, and taking into account different aspects and views(Del Felice et al 2017:9).

The term “counter narratives” is used mainly in the context of the fight against extremism and terrorism and emphasizes the need to dismantle and devaluate violent narratives (Gelber 2002:135) that may seem appealing, especially to young people: “*Extreme right-wing discourse is attractive for young people, especially adolescent boys who are attracted to the "macho" profile promoted by these political ideologies*” (JS).

The use of the term “alternative narratives” emphasizes the importance of presenting different aspects of an issue and highlights positive alternatives that extend beyond the negative picture painted by the narratives they seek to address. The distinction between the two terms is often indistinguishable in practice, as a counter narrative presupposes or indirectly refers to an alternative narrative.

In short, *the “counter narrative” creates a message with coherent content that attempts to deconstruct the content of hate discourse, while the “alternative narrative” focuses on different -and often positive- parameters of an issue, trying to create an alternative approach compared to the one presented by hate discourse*. Human rights -should- constitute the basis of the narratives combating hate discourse. Human rights-based narratives play an important role in the emancipation strategies of young people, especially those who have experienced hate discourse closely, either as targets or perpetrators (Del Felice et al. 2017:10).

Regarding the tackling of direct/straightforward hate discourse, in the framework of the international program “Antigone”, the participants, after discussing cases of direct/straightforward hate speech expression against immigrants, named ***six different ways of dealing with hate discourse***.

|  |
| --- |
| **Different ways of dealing with Hate Discourse*** Through the organization of actions for and with the public not typically involved in such initiatives (dissemination to the community).
* Through cooperation with political institutions and competent authorities.
* Focusing on the term itself (used in the context of hate discourse expression), in order to highlight the political dimension of the issue and the current political situation.
* By creating a special dictionary that will be shared and used in schools.
* Through education (on human rights and the obligation to respect them).
* Through media education.
 |

The above actions can be implemented both through civil society actors and at the individual level (Antigone 2019:23).

Examples of indirect hate discourse were then discussed as part of the same program. In the first example, an immigrant kindergarten student experienced isolation from her classmates, due to false rumors that her mother worked as a prostitute. The rumors were circulated by parents and also affected the student's teacher. In combination with a media focus -at that time- on women of African descent working as prostitutes in Greece, the child had been significantly affected. The second case involved a young second-generation immigrant child of African parents, who began to have problems due to the behavior of his classmates and friends, triggered by the growing intensity of the immigration issue. As a result, the child became socially isolated and became increasingly active on the Internet, which made him prone to aggressive behaviors by strangers. In the end, the participants in the workshop named four possible ways of dealing with indirect hate discourse:

1. **Education**: educational actions in schools and workplaces teaching acceptance of diversity. Emotional intelligence training, cultivating a “team philosophy”, discussing hate discourse during school hours and open and ‘unified’ parent councils.

2. **Children**: emphasizing the importance of children talking about hate discourse and informing adults when they become victims of hate discourse.

3. **Public figures**: informing and raising awareness in the general population in order to reduce ‘neutrality’, informing and supporting teachers to help them, in turn, support children on the subject of hate discourse.

4. **Mainstream Media**: public television should focus on groups of people who are willing to speak out as victims of various forms of violence. The purpose of such a television program is to encourage those who have been victims of hate discourse and subsequent acts of violence to speak out.

5. **Arts**: discussing about hate discourse that diffuses through various art forms (Antigone 2019:24).

As part of the same program, the subgroups of online hate discourse were given three examples of hate discourse on twitter for discussion. The moderator of the discussion explained the terminology around hate discourse and then stressed the fact that it is almost impossible to control it.

The three examples shared by the moderators opened up a discussion about the ways in which social media users can utilize them to respond to hate discourse or to avoid falling victim to it. Education emerged -once again- as the main tool that can help combat such a phenomenon.

When asked if social media users should respond to online hate directed towards them (either verbally or by deleting such comments), participants agreed that this is a very sensitive issue, as deleting, for example, a comment can be considered censorship. *The participants suggested the creation of an algorithm or instrument that can be applied by social media companies to ban or delete specific words or posts that someone may post or upload.* Finally, participants pointed out that reporting online hate discourse to the Cyber Crime Unit of the Hellenic Police is helpful, as they have the tools to monitor hate discourse exponents/perpetrators and to take action if necessary (Antigone 2019:26).

The main conclusion that emerged from this program is that hate discourse can be addressed through formal and non-formal education and, therefore, the competent bodies should focus their attention on educational actions, which will raise awareness among students. Finally, one of the main ways to raise awareness about hate discourse, both online and offline, is through collective action with the participation of both official (government) and civil society organizations (Antigone 2019:26).

|  |
| --- |
| Hate discourse can be addressed through formal and non-formal education. Therefore, the competent bodies should focus their attention on educational actions, aiming to raise awareness among students. |

It follows from the above that those who participated in the “Antigone” program eventually came to the same conclusion: hate discourse can be addressed and combated through education. Not just any education, but an education and education system that puts human rights at its core. Complementarily and in combination with proper education, public awareness campaigns can be useful tools in fighting this phenomenon (Antigone 2019:28).

# Some practical advice

In addition to dealing with hate discourse when it emerges, it is always good to try -to the best of our abilities- to prevent its manifestation. In the context of participating in public speech, the prevention of hate discourse becomes even more important. Based on this logic, we present some practical tips that have been included in the “Guide for Journalists” of the SOpHiSM program.

Given the fact that, on the Internet, journalists are not the only news reporters, we believe that the following tips are useful for anyone who expresses themselves -systematically or not- within the online public sphere. The following tips, in addition to hate discourse prevention practices, are also potential parameters for identifying the existence of hate discourse in messages that do not follow even some of these tips.

1. When writing a message, it is good to take into account the element of diversity, i.e. cultural, racial, ethnic, religious, social, sexual, in order to promote mutual understanding between the different groups of the population.
2. Reference to the ethnic and/or racial origin of individuals should be avoided unless it is an integral part of the information. Especially in cases of crimes or offenses, no reference should be made to the element of national origin, religion of the suspect or suspects if this is not a necessary component of the incident.
3. Large periods of economic crisis and high unemployment create fertile ground for the creation and spread of feelings of xenophobia, racism, and hostility towards immigrants. This requires all of us to be especially careful when handling information about immigrants, especially asylum seekers, refugees, victims of trafficking, or other vulnerable social groups. At the same time, it is good to try, in our own way, to give a voice to these people or, if we do not have the means to do so, to become their own voice.
4. We must always respect the presumption of innocence and not disclose information relating to the private life of individuals involved in criminal acts. Everyone has personal data (and is entitled to their protection), whether they are indigenous, immigrants, or belonging to social or population minorities.
5. When writing the message, it is better to avoid leveling, absolute, derogatory expressions and characterizations, mainly against people who belong to vulnerable groups and do not have the ability to defend themselves.
6. The identities of asylum seekers, refugees, victims of human trafficking, and even immigrants who choose to speak out in public should not be disclosed, as they or their relatives may face retaliation from their country's authorities or by non-governmental and even criminal organizations. For the same reason, photos showing asylum seekers and victims of exploitation should not be displayed. If circumstances require publication/display of such information, this should be done in a way that does not render the persons identifiable.
7. In messages concerning immigrants, it is preferable to avoid the use of standard file photos, e.g. women in veils, headscarves, men in prayer position. Not all immigrants are Muslims or religious Muslims, while stereotypes may create negative associations (prejudices).
8. When the message concerns cases of women who are victims of human trafficking, in no case should it be implied that these women are extradited on their own free will. On the contrary, it must be emphasized that these women are being inhumanely exploited.
9. The messages should not be intended to impress or create feelings of fear, panic, or anxiety regarding the effects of immigration, unless there is serious and well-founded evidence for the allegations made.
10. When circumstances require the publication of statements with adverse content directed against individuals belonging to minorities, the right to respond must be granted at all costs.
11. When possible, it is advisable to seek the advice of experts or organizations that are well informed about the issues of immigrants, refugees, asylum seekers, victims of human trafficking, sexual orientation, religious minorities, the disabled, etc. in order to convey clear and comprehensive information to the public.

The following section contains indicative references to the correct use of words and terminology:

1. The term preferred by international human rights bodies for immigrants entering or residing in a country without legal documents is “illegal immigrants” or “undocumented immigrants”.
2. An asylum seeker is a person who is outside their home country and who has applied to be recognized as a refugee or to receive some other kind of international protection. The asylum seeker, regardless of whether they entered the country without legal documents or in any other irregular manner, has the right to remain in the country, which is in turn responsible for examining the application request until the final decision has been taken.
3. A refugee is a person who has been granted this status under the 1951 Geneva Convention, which has been endorsed by 148 countries. According to Article 1 of the Treaty, a refugee is a person who “owing to wellfounded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country”. Refugee status can be granted to people who meet the above conditions.
4. Beneficiary of Subsidiary Protection is a status granted to asylum seekers who do not face the risk of personal persecution in their home country under the Geneva Convention, but cannot return to their homeland because they risk serious harm -torture, death penalty, inhuman or humiliating treatment, serious threat to their lives- due to conflict or widespread violence.
5. A victim of trafficking is a person who is being exploited, whether it be for work, sexually or physically, such as the removal of organs, by other people who exercise complete control over them. The victim of human trafficking has either not given their consent to be transferred to another country, or even if they have given their consent, their consent is annulled due to coercion and deception suffered by the actions of the traffickers or the abuse they have suffered or with which they were threatened.
6. An immigrant is a person who chooses voluntarily to look for work and better living conditions in another country. Unlike a refugee, the immigrant's life and liberty are not in danger and they can return to their home country if they so wish. Nevertheless, we must keep in mind the fact that people usually do not leave their homes and families ‘light-heartedly’ and behind the immigrants' choice to leave are often extremely difficult and adverse -e.g. based on economic criteria- living conditions (e.g. conditions of absolute poverty).
7. An illegal immigrant or “undocumented immigrant” is someone who enters a country without being checked at the border, or someone who remains in a country after the end of their legal stay, or someone who remains in a country after an extradition or deportation order has been issued.
8. When referring to LGBT+ people, words such as “immoral”, “sinners”, “lustful”, “filthy”, “damned” are not used. Typically, the use of these terms has a religious background and are used mainly by priests who do not accept diversity in relation to sexual orientation.
9. The use of nouns such as “divergent”, “disturbed”, “dysfunctional”, “perverted”, “sick”, “abnormal”, “destructive” is intended to present LGBT+ people as subhumans, mentally ill, or a risk to society. It is noted that the perception that LGBT+ people suffer from a mental disorder has been debunked by scientific research.
10. It is emphasized that the use of words with offensive content, words with a sexist background aimed at ridiculing, humiliating, and abusing people of different sexual orientation is considered hate rhetoric. When referring to LGBT+ people, the terminology promoted by the European Commission against racism and intolerance should be used instead of abusive designations, such as “faggot”, “fag”, “sissy”, “pantywaist”, “fairy”, “poofter”, “bent”, “shemale”, “tranny” etc.
11. LGBT+ people flirt, they do not “deceive”, “trap”, “mock”, “pretend”, “act”, “fake it for publicity”, “disguise”.
12. Homosexual/Gay is the man who is sexually and/or emotionally attracted to men. A lesbian is a woman who is sexually and/or emotionally attracted to women.
13. Bisexual is a person who feels sexual and/or emotional attraction for more than one sex.
14. Transgender/Trans people are people who have a gender identity, which differs from the gender assigned to them at birth.
15. Asexual is a person who experiences no or very little sexual attraction. Intersex are people who are born with gender characteristics (such as chromosomes, genitals, and/or hormonal structure) that do not belong strictly to male or female categories, or belong to both at the same time. Pansexual/Omnisexual are people who have a romantic, sexual or emotional desire for people regardless of gender identity.
16. Homophobia is fear, unwarranted anger, intolerance, and/or hatred towards homosexuality
17. Sexism is any action, gesture, visual presentation, oral or written discourse, practice, or behavior based on the idea that an individual or group of individuals is superior because of their gender, taking place in public or in the private sphere, either online or offline.
18. Black people are referred to as “black” or, if they are American citizens, “African-American”. The description “colored” is considered wrong, if not racist since all people have color.

# 3.4 Summary

In this section we have emphasized the importance of different methods of addressing and dealing with hate discourse, starting from the legal framework that exists at national and international level and noting that the legal arsenal of institutions that wish to deal effectively with hate discourse has several weapons. Of course, any use (or not) of this arsenal is at the disposal of the various executive and legislative authorities. In any case, the legal/criminal treatment of this issue is merely an attempt to ‘cure’ the symptoms and not ‘prevent’ the emergence of hate discourse. At the same time, we highlighted the importance of continuous formal and informal education for effectively addressing and dealing with hate discourse, which can focus on combating the causes of the phenomenon.

We must emphasize that even a coordinated effort to tackle hate discourse by any means available does not guarantee the success of the project. The ‘uprooting’ of hate discourse from societies is a long process with a continuous ‘intensity’ of actions that requires dedication, mobilization, and collective efforts of as much of the society as possible.

# Instead of a conclusion

An the end of this educational material we will make a brief summary of the logic that governs it. First of all, in what has to do with the definition and description of (online) hate discourse, this text adopts a social constructionist logic in conjunction with the critical speech analysis approach, in order to define this phenomenon as a predominantly social phenomenon manifested through discourse, i.e. the use of every possible means of communication (language, images, symbols, etc.), in order to spread messages of hatred primarily -but not exclusively- against specific and based on social-ideological criteria marginalized social groups.

We, then, referred to the attempts to legally regulate (online) hate discourse incidents, while presenting some basic legal texts that define and aim to address (online) hate discourse. At this point we noted that legal texts and attempts to regulate the phenomenon are necessary conditions in the fight against hate discourse and have an important role to play, primarily in punishing the perpetrators, and secondarily in reducing the phenomenon.

Nevertheless, *the deep social-ideological roots of (online) hate discourse require a painful and constant struggle that should be carried out through continuous education programs (formal and informal) and public efforts to spread information and raise awareness, through the closest and most well-coordinated possible actions of the actors involved in this field*. Another ‘stepping stone’ in this effort is the SoPHiSM program, which focuses its actions on the social research-education-dissemination triptych.

This triptych is not randomly proposed, as one should conduct a thorough research on the different forms that hate discourse can present, both online and offline, as well as in what has to do with the causes of its manifestation. Only by fully understanding the causes of hate discourse at both national and supranational levels can we plan the appropriate actions to address it. In addition, it is equally important to raise public awareness regarding discrimination issues. This can only be accomplished through the organization of relevant events, the adoption of a responsible attitude by the media and, of course, through an agenda that promotes diversity and tolerance at all levels of education (Assimakopoulos, Baiden and Millar 2017:89).

This educational material does not offer ready-made alternatives to hate discourse for obvious reasons: (online) hate discourse can take many different forms and any response should take into account specific factors, including the text and the social context, and, of course, language. Besides, human rights activists and campaign leaders must have the autonomy and confidence to develop their own responses to hate discourse, and not necessarily reproduce someone else's -ready-made- responses, even if that ‘someone else’ is the Council of Europe (Del Felice et al. 2017:16).

Hate discourse as a social phenomenon cannot be approached as an isolated problem or treated with an individualistic logic. It must be seen as an issue concerning our collective, democratic, social (co)existence. In this context, educational and emancipatory strategies, using counter and alternative (to hate discourse) narratives, are critical coping tools. (Online) hate discourse is not a problem that concerns only those immediately affected by it, but instead, to some degree it affects and concerns us all (Del Felice et al. 2017:43; Gelber 2002:11). By this logic we stand against it and aspire to contribute -albeit to a minimum- through the SoPHiSM program to the creation of a more open society.

# References

Acula D. D., Oblan L. A. C., Pedroso T. B., Riosa K. J. V., Tolibas M. A. R. (2018). Implementing Fact-Checking in Journalistic Articles Shared on Social Media in the Philippines Using Knowledge Graphs, 2018 3rd International Conference on Computer and Communication Systems, 462-466.

Alkiviadou N. (2018). Regulating hatred: Of devils and demons? *International Journal of Discrimination and the Law*, (18)4, 218-236.

Antigone (2019). *Words that Lead to Hate: Hate Speech in Greece during 2018*. Antigone Information and Documentation Centre on Racism, Ecology, Peace and Non-violence.

Assimakopoulos S., Baiden F. H., και Millar S. (2017). *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*. Switzerland: Springer Open.

Benesch S. (2012). Words as Weapons. *World Policy Journal*, 29(1), 7-12.

Blakemore M., Longhorn R. (2001). Communicating Information about the World Trade Center Disaster: Ripples, Reverberations, and Repercussions. First Monday, volume 6(12). Διαθέσιμο στη διεύθυνση: https://journals.uic.edu/ojs/index.php/fm/article/view/907/816 (τελευταία προσπέλαση 11 Αυγούστου 2020).

Blommaert J., Bulcaen C. (2000). Critical Discourse Analysis. *Annual Review of Anthropology*, 29, 447-466.

Brown A. (2017). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297-326.

Brown A. (2020). *Models of Governance of Online Hate Speech*. Council of Europe: Council of Europe Publications.

Brown A. and Sinclair A. (2019). *The Politics of Hate Speech Laws*. London: Routledge.

Buyse A. (2014). Words of Violence: “Fear Speech”, or How Violent Conflict Escalation Relates to the Freedom of Expression. *Human Rights Quarterly*, 36(4), 779-797.

Castells M. (2000). *The Rise of the Network Society*. Oxford: Blackwell Publishing.

Chomsky Ν. (2003). *Tα ΜΜΕ ως 'Οργανο Κοινωνικού Ελέγχου και Επιβολής*. Αθήνα: Ελεύθερος Τύπος.

Cochliou D. (2020). *SOpHiSM Comprehensive Report*. SOpHiSM: A Response to Online Hate Speech through the Enhancement of high-quality Professional and Citizen Journalism.

Cohen-Almagor R. (2012). Is Law Appropriate to Regulate Hateful and Racist Speech? The Israeli Experience. *Israel Studies Review*, 27(2), 41-64.

Cohen-Almagor, R. (2015). *Confronting the Internet’s Dark Side: Moral and Social Responsibility on the Free Highway*. Cambridge: Cambridge University Press.

Council of Europe (1961). *European Social Charter (Ευρωπαϊκός Κοινωνικός Χάρτης)*. Διαθέσιμο στη διεύθυνση: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/035. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Council of Europe (1984). *Council of Europe Framework Convention for the Protection of National Minorities (Σύμβαση-πλαίσιο του Ευρωπαϊκού Συμοβυλίου για την προστασία των εθνικών μειονοτήτων)*. Διαθέσιμο στη διεύθυνση: https://www.coe.int/en/web/minorities/at-a-glance. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Council of Europe (2000). *Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms*. Διαθέσιμο στη διεύθυνση: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/177. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Council of Europe (2015). *Convention on Cybercrime: Protocol on Xenophobia and Racism (Πρωτόκολλο για την Ξενοφοβία και τον Ρατσισμό)*. Διαθέσιμο στη διεύθυνση: https://edoc.coe.int/en/cybercrime/6559-convention-on-cybercrime-protocol-on-xenophobia-and-racism.html. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Council of Europe. Committee of Ministers (1997). *Recommendation R (97) 21 on the Media and the promotion of a culture of tolerance (Σύσταση σχετικά με τα μέσα ενημέρωσης και την προώθηση μιας νοοτροπίας ανεκτικότητας)*. Διαθέσιμο στη διεύθυνση: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168050513b. Τελευταία προσπέλαση 12 Αυγούστου 2020.

De Latour A., Perger N., Salaj R., Tocchi C., Otero-Viejo P. (2017). *We Can! Taking Action against Hate Speech through Counter and Alternative Narratives*. Revised edition, Council of Europe.

Del Felice C., Ettema M., de Latour A., Perger N., Salaj R., Tocchi C., Viejo Otero P., Gomes R. (2017). *We Can! Taking Action against Hate Speech through Counter and Alternative Narratives*. Council of Europe: Council of Europe Publications.

European Commission (2014). *Special Eurobarometer 415: Public Perceptions of Science, Research and Innovation*. European Commission, Directorate-General for Research & Innovation (DG RTD).

European Commission (2015). *Special Eurobarometer 437: Discrimination in the EU in 2015*. European Commission, Directorate-General for Research & Innovation (DG RTD).

European Commission (2016). *Autumn 2016 Standard Eurobarometer: Immigration and terrorism continue to be seen as the most important issues facing the EU*. European Commission press release. Διαθέσιμο στη διεύθυνση: http://europa.eu/rapid/press-release\_IP-16-4493\_en.htm. Τελευταία προσπέλαση 12 Αυγούστου 2020.

European Commission (2016). *Code of conduct on countering illegal hate speech online.*

European Commission against Racism and Intolerance (2016). *ECRI General Policy Recommendation 15*. European Commission against Racism and Intolerance- Council of Europe.

European Commission document. Διαθέσιμο στη διεύθυνση: http://ec.europa.eu/justice/fundamental-ights/files/hate\_speech\_code\_of\_conduct\_en.pdf. Τελευταία προσπέλαση 12 Αυγούστου 2020.

European Court of Human Rights (2010). *Ευρωπαϊκή Σύμβαση Δικαιωμάτων του Ανθρώπου (European Convention on Human Rights)*. Strasbourg: European Court of Human Rights.

Foucault M. (1977). The archaeology of knowledge. London, England: Tavistock.

Foucault M. (1979). Discipline and punish: The birth of the prison. New York, NY: Vintage.

Gagliardone I., Danit G., Alves T., Martinez G. (2015). *Countering online hate speech. Unesco Series on Internet Freedom*. Διαθέσιμο στη διεύθυνση: http://unesdoc.unesco.org/images/0023/002332/233231e.pdf. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Gale, P. (2004). The refugee crisis and fear: Populist politics and media discourse, *Journal of Sociology,* 40(4): pp. 321-340.

Gelber K. (2002). *Speaking Back: The Free Speech versus Hate Speech Debate*. Amsterdam/Philadelphia: Jonh Benjamins Publishing Company.

Ging D., Siapera E. (2019). *Introduction*. Στο Ging D., Siapera E. (επιμ.) Gender Hate Online. Understanding the New Anti-feminism. Cham, Switzerland: Palgrave McMillan.

Hawdon J., Oksanen A., Rasanen P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, 38, 254-266.

Heinze E. (2016). *Hate Speech and Democratic Citizenship*. Oxford: Oxford University Press.

In-Cyprus (2020). *Racism Watch: The normalization of racist speech in Cypriot society must end*. Διαθέσιμο στη διεύθυνση: https://in-cyprus.philenews.com/racism-watch-the-normalization-of-racist-speech-in-cypriot-society-must-end/?fbclid=IwAR2GJ4TIVzMsfffs3T921w8MMT0IWVJJimxP7KapqkrJlK7goPfb5H7vks0. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Jakubowicz A, Dunn K., Mason G., Paradies Y., Bliuc A. M., Bahfen N., Oboler A., Atie R., Connelly K. (2017). Cyber Racism and Community Resilience. Strategies for Combatting Online Race Hate. Cham, Switzerland: Palgrave McMillan.

Jakubowicz A., Dunn K., Mason G., Paradies Y., Bliuc A. M., Bahfen N., Oboler A., Atie R., Connelly K. (2017). Cyber Racism and Community Resilience: Strategies for combatting Online Race Hate. Switzerlad: Palgrave Macmillan.

Kalsnes B. (2018). Deciding what’s true: The rise of political fact-checking in American journalism, *Digital Journalism*, 6(5), 670-672.

La Macchia S. T., Louis W. R. (2016). Crowd behaviour and collective action*.* Στο McKeown, S., Haji R., Ferguson N. (επιμ.). *Peace psychology book series. Understanding peace and conflict through social identity theory: Contemporary global perspectives,* 89-104.

Laclau E., Mouffe C. (1985). *Hegemony and Socialist Strategy. Towards a Radical Democratic Politics***.** 2nd edition. London, New York: Verso.

Lim C. (2018). Checking how fact-checkers check. *Research and Politics*, 1-7.

London T. M., Crundwell J., Bock Eastley M., Santiago N., Jenkins J. (2020). Finding Effective Moderation practices on Twitch. Στο Reyman J., Sparby E. M. (επιμ.) *Digital Ethics: Rhetoric and Responsibility in Online Aggrsssion.* New York and London: Routledge, 51-68.

Mansell R., Steinmueller W. E. (2002). *Mobilizing the Information Society*. Oxford: Oxford University Press.

Mantzarlis A. (2018). ‘Fact-checking 101’. Στο C. Ireton and J. Posetti (επιμ.) *Journalism, Fake News and Disinformation. Handbook for Journalism Education and Training*. Paris: United Nations Educational, Scientific and Cultural Organization, 85-100.

McNair B. (2006). Cultural Chaos. Journalism and Power in a Globalized World. London: Routledge.

Miskolci J., Kovacova L., Rigova E. (2018). Countering Hate Speech on Facebook: The Case of the Roma Minority in Slovakia. Social Science Computer Review, XX(X), 1-19.

Mollen A. (2018). *Digital Spaces of Civic Communication. The Practices and Interfaces of Online Commenting*. Wiesbaden: Springer VS.

Mollen A. (2020). *Digital Spaces of Civic Communication*. Wiesbaden, Germany: Springer VS.

Oksanen A., Hawdon J., Holkeri E., Nasi M., Rasanen P. (2014). Exposure to Online Hate among Young Social Media Users. *Soul of Society: A Focus on Lives of Children and Youth-Sociological Studies of Children and Youth*, 18, 253-273.

Porter J. E. (2020). Interacting with Friend, Enemies and Strangers. Στο Reyman J., Sparby E. M. (επιμ.) *Digital Ethics: Rhetoric and Responsibility in Online Aggrsssion.* New York and London: Routledge, xvi-xxii.

Poster M. (1995). *The Second Media Age*. Cambridge: Polity Press.

Poulakidakos S. (2018). The discourse on immigration in the Greek public sphere: Between securitization and humanitarianism. *Science and Society*, 37, Winter 2018, 109-138.

Poulakidakos S., Veneti A. (2016). Political Communication and Twitter in Greece: Jumping on the bandwagon or an enhancement of the political dialogue? Στο Deželan T. and Vobic I. (επιμ.) *(R)evolutionizing Political Communication through Social Media*. Hershey PA: IGI Global, 119-146.

Poulakidakos S., Veneti A., Fangonikolopoulos Ch. (2018). Post-truth, propaganda and the transformation of the spiral of silence. *International Journal of Media and Cultural Politics*, 14(3), pp. 367-382.

Reyman J., Sparby E. (2020). Introduction: Toward an Ethic of Responsibility in Digital Aggression. Στο Reyman J., Sparby E. M. (επιμ.) *Digital Ethics: Rhetoric and Responsibility in Online Aggrsssion.* New York and London: Routledge, 1-14.

Royzman E. B., McCauley C., Rozin P. (2005). From Plato to Putnam: Four Ways to Think About Hate. Στο R. J. Sternberg (επιμ.). *The psychology of hate*, 3-35.

Schmidt S. J. (2001). *Media societies: Fiction Machines*. Στο Grant C.B., McLaughlin D. (επιμ.) Language – meaning – social construction. Interdisciplinary studies. New York: Rodopi.

Sobieraj S. (2019). *Disinformation, Democracy, and the Social Costs of Identity-Based Attacks Online*. Διαθέσιμο στο https://mediawell.ssrc.org/expert-reflections/disinformation-democracy-and-the-social-costs-of-identity-based-attacks-online/. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Special Rapporteur on the promotion and protection of the freedom of opinion and expression (2019). *Report of the Special Rapporteur on the promotion and protection of the freedom of opinion and expression*. United Nations.

Sponholz L., Christofoletti R. (2019). From preachers to comedians: Ideal types of hate speakers in Brazil. *Global Media and Communication*, 15(1), 67-84.

Strossen N. (2018). *HATE: Why we should resist it with Free Speech, not Censorship*. Oxford: Oxford University Press.

Taguieff, P. A. (1993). From Race to Culture: The New Right’s View of European Identity, *Telos* (98-99): pp. 99-125.

Terranova T. (2004). Network Culture, Politics for the Information Age. London: Pluto Press.

The Committee of Ministers (1997). *Recommendation No. R (97) 20 of the Committee of Ministers to member states on “hate speech” (Sύσταση υπ’ αριθμό R (97) 20 προς στα κράτη μέλη σχετικά με το Λόγο μίσους)*. Διαθέσιμο στη διεύθυνση: https://www.coe.int/en/web/freedom-expression/committee-of-ministers-adopted-texts/-/asset\_publisher/aDXmrol0vvsU/content/recommendation-no-r-97-20-of-the-committee-of-ministers-to-member-states-on-hate-speech-?\_101\_INSTANCE\_aDXmrol0vvsU\_viewMode=view. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Thompson J. (1999). *Νεωτερικότητα και Μέσα Επικοινωνίας*. Β’ έκδοση, Αθήνα: Παπαζήσης.

Waldron J. (2012). *The Harm in Hate Speech*. Cambridge Massachussets, London England: Harvard University Press.

Wiggins R. (2001). The effects of September 11th on the leading search engine. First Monday, 7(10). Διαθέσιμο στη διεύθυνση: http://web.mit.edu/6.033/2002/wwwdocs/papers/9-11.html. Tελευταία προσπέλαση 11 Αυγούστου 2020.

Wodak R. (1995). Critical linguistics and critical discourse analysis. Στο Verschueren J., Ostman J.O., Blommaert J. (eds.) *Handbook of Pragmatics 1995*. Amsterdam: Benjamins, 204-210.

Δεμερτζής Ν. (2002). *Πολιτική Επικοινωνία, Διακινδύνευση, Δημοσιότητα, Διαδίκτυο.* Αθήνα: Παπαζήσης.

*Διεθνής Συνθήκη Ατομικών και Πολιτικών Δικαιωμάτων (I*nternational Covenant on Civil and Political Rights*-ICCPR)* (1966). Διαθέσιμο στη διεύθυνση: https://www.refworld.org/cgi-bin/texis/vtx/rwmain/opendocpdf.pdf?reldoc=y&docid=4bd686e52. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Δίκτυο Καταγραφής Περιστατικών Ρατσιστικής Βίας (2018), Ετήσια Έκθεση 2017. Διαθέσιμο στη διεύθυνση: http://rvrn.org/wp-content/uploads/2018/03/Report\_2017gr.pdf. Τελευταία προσπέλαση 12 Αυγούστου 2020.

*Οικουμενική Διακήρυξη των Δικαιωμάτων του Ανθρώπου (Universal Declaration of Human Rights-UDHR)* (1948). Διαθέσιμο στο https://www.amnesty.gr/universal-declaration-of-human-rights. Τελευταία προσπέλαση 12 Αυγούστου 2020.

Πλειός Γ. (2011). *Η Κοινωνία της Ενημέρωσης*. Αθήνα: Καστανιώτης.

Πουλακιδάκος Σ. (2014). *Προπαγάνδα και Δημόσιος Λόγος. Η παρουσίαση του Μνημονίου από τα Ελληνικά ΜΜΕ*. Αθήνα: Εκδόσεις DaVinci.

1. In the framework of this program, semi-structured interviews were conducted with journalism students (JS), young journalists (YJ), and “citizen journalists” (CJ), as well as focus groups including journalism students, young journalists, and citizen journalists (FGJ) on the one hand, and human rights defenders (FGHR) on the other, both in Greece and in Cyprus, regarding hate discourse on the Internet. Certain characteristic excerpts derived from this research are presented in this educational material, with the aforementioned abbreviations in parentheses revealing the ‘origin’ of the characteristic excerpt quoted. [↑](#footnote-ref-1)
2. A)Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation - *Official Journal L 303 , 02/12/2000 P. 0016 - 0022*.

B) Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin - *Official Journal L 180 , 19/07/2000 P. 0022 - 0026.*

C) Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

D) Manual for gender mainstreaming: Employment, social inclusion and social protection policies: https://eige.europa.eu/node/636. [↑](#footnote-ref-2)
3. Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, OJ L 328, 6.12.2008. [↑](#footnote-ref-3)
4. The EU Code of conduct on countering illegal hate speech online, 30 June 2016. [↑](#footnote-ref-4)
5. The evaluation reports can be found at: <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> [↑](#footnote-ref-5)