HATE SPEECH IN CYPRUS

Arianna Tortelli, Project Manager - Research & Development Division
Lorenzo Italiano, Intern - Research & Development Division

Contents

HOPE For Children

HATE TRACKERS BEYOND BORDERS

Agenzia Italiana per la Gioventù

Co-funded by the European Union

INTRODUCTION

Over the past thirty years, the phenomena of hate speech, violence, and discrimination have gradually increased, to the point of becoming the focus of public debate, and with it, these topics have steadily begun to be discussed by academics, scholars, and have been addressed by institutions and international organizations. The main cause ascertained by the vast majority has been the advent and spread of social media.

This report introduces the concept of hate speech, derogatory language, and its characteristics. The first part of the report will analyze the phenomena of hate speech and discrimination in a wider context where they are embedded. The report will also explore the debate about freedom of expression between the U.S. legal approach and the EU legislative framework ruling it. Most studies conducted on hate speech consider European and U.S. legal contexts, yet the major social media platforms (Meta, Reddit, X, YouTube, Google) are owned by U.S. companies, except for TikTok, which is owned by Zhang Yiming's Chinese tech giant ByteDance.

The second section of the report will focus on the definition of hate speech, its effects, the Pyramid of Hate set out by the Council of Europe, and the risks of online activism. It will detail the consequences of hate speech, introducing also the Psychological First Aid Toolkit developed by "Hope For Children" CRC Policy Center within the framework of the Hate Trackers Beyond Borders project. Finally, there will be a dedicated analysis of the role of social media concerning hate speech, explaining the concepts of media bubbles, while also giving an insight into how social media and their algorithms work.

Finally, the third and fourth sections will provide an overview of the socio-political and legal context of Cyprus and local services to support victims of hate speech.

This report aims to deepen knowledge and understanding of hate speech, particularly in Cyprus, and provide the essential tools and information to recognize its impact on society.

FREEDOM OF EXPRESSION (U.S. vs EU LEGAL FRAMEWORK)

Hate speech, derogatory language, public speeches, and propaganda aimed at spreading and promoting hatred and violence have emerged as issues in contemporary society. Only in the last three decades, however, the number of studies on the subject has increased significantly. By the end of the 20th century, the phenomenon of hatred, hate speech, and discrimination, and their dangerous effects, gained more attention and were included in the public debate, becoming the subject of interest in social, legislative, and policy studies. One of the major causes identified by most scholars, associations, international organizations, and institutions has been the emergence and spread of social media.

Geographical studies regarding hate speech, racism, and derogatory language on social media, have mainly focused on the context of the United States with a 44.23% rate globally, while Europe ranks second with more than 26% (Matamoros-Fernandez and Farkas 2020). Authors such as Paz M. A., Montero-Diaz J., and Moreno-Delgado A., have shown that the entire number of legal studies on hate speech conducted in the period between 1975 and 2019 focused largely on the United States (with a percentage of about 50 % of the total), while Europe accounted for more than a third of the total. Therefore, since nearly 80 % of the studies carried out on hate speech were conducted by analyzing the contexts of the United States and the European Union, the debate on the topic revolves around these two legal approaches.

The First Amendment of the U.S. Constitution claims that "Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof; or abridging the freedom of speech, or the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances". With the First Amendment the U.S. Constitution enshrines and aims to protect full freedom of expression, and the only point in which speech may be subject to limitation is when it enters within the definition of "fighting words" as words uttered directed at the person of the listener that would tend to provoke an act of violence (Chaplinsky v. New Hampshire, 315 U.S. 568, 1942). This means the ideological strand carried by the Supreme Court justices has been to protect full freedom of thought and speech. A clear statement explicative of this philosophy of thought was made by the judges of the Supreme Court during some of its historic sentences: "one man's vulgarity is another man's lyricism" (Cohen v. California, 403 U.S. 15, 1971) or "under the First Amendment we must protect the ideas we detest, otherwise sooner or later they will prohibit us from expressing the ones we love" (Communist Party of the United States v. SACB, 367 U.S. 1, 1961). The last quoted sentence is relevant because it constitutes one of the earliest decisions of the Supreme Court regarding incitement to hatred. What happened is that the National Socialist Party of America decided to conduct a demonstration in Skokie, a mainly Jewish neighbourhood of Chicago inhabited by approximately 7000 people who had survived the Holocaust. An injunction was signed to stop the rally, which was overturned by the Illinois Supreme Court as being contrary to the First Amendment, but the Court ruled that the party had the right to demonstrate, stating that "the public expression of ideas cannot be prohibited simply because the ideas themselves

are offensive to some of their hearers" (U.S. Supreme Court 1978).

While the U.S. Supreme Court throughout the years has ruled that freedom of expression is fully protected until it can directly and concretely provoke an act of violence, in Europe the legal approach towards freedom of expression and derogatory language is different. In the EU legal framework, freedom of expression is not an absolute right, but it can be subject to formalities, restrictions, and conditions when it affects other core democratic values.

The European Convention on Human Rights (ECHR), ratified by Cyprus in 1962, balances freedom of expression with other core values of democracy, equality, non-discrimination, human dignity, and human integrity. For instance, the Article 9 of ECHR states that freedom to manifest one's religion or belief can be, in certain cases, subject to limitation "[…] in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others". Also, Article 10 of ECHR states that, despite everyone must be guaranteed the right of freedom of expression and the freedom to hold opinions, in some cases the exercise of this right can be set to restrictions or penalties when it affects other key principles necessary to the democratic participation. Other core rights deal with the prohibition of discrimination (art. 14), and prohibition of abuse of rights (art. 17). Thus, freedom of expression (Art.10 CHR) is not an absolute right, it is rather a right that contains many limitation mechanisms and barriers. That is why Article 10 of the ECHR is interpreted very broadly so that each EU Member State can adapt it to its jurisdiction, needs, culture, and national security standards (Mendel, rm.coe.int/16806f5bb3). Especially when an overriding interest is at stake, like one of public security, the freedom of expression is limited, and the other interest prevails.

To show proof of the different approach of the Council of Europe regarding freedom of expression, there is a relevant sentence of the European Court of Human Rights (ECtHR), titled "M'Bala M'Bala v. France (2003)". In 2008, Dieudonné M'Bala M'Bala, a French comedian, invited the academic Robert Faurisson, who had denied the existence of gas chambers in concentration camps, on stage during his show and had him applauded by the crowd. The academic had already received several convictions for his denialist theories. The showman presented him with an award for "nonchalance and insolence", and had it given to him by an actor dressed in striped pyjamas and the Star of David.

The comedian was convicted in 2009 by a court of first instance in Paris. The conviction was confirmed two years later by the Paris Court of Appeal and the appeal rejected by the Court of Cassation. In April 2013, the comedian appealed to the ECtHR claiming that the sentence had violated the freedom of expression enshrined in Article 10 of the ECHR. The court held that the comedian's TV show went beyond satire and did not have the protection of Article 10 because it was judged to be a manifestation of anti-Semitic hatred and Holocaust denial. While satirical speech is protected, the ideology expressed in the show, by inviting a Holocaust denier on stage, was contrary to the values of justice and peace expressed by the ECHR.

EU agreements with private IT companies on the management of hate speech

Having explained the difference between the two different models, this section aimsto explore the agreements made with private IT companies on hate speech.

An important document is a soft-law document adopted by the European Commission in 2016, the "Code of conduct on countering illegal hate speech online", which provides for the negotiation of a set of rules with private IT companies, among the most important being Facebook, Microsoft, Snapchat (from 2018), X (former Twitter), Youtube, Linkedin, Instagram (from 2019) and TikTok (added in 2021), giving them incentives to comply with the agreed rules. The Code calls on companies "to have rules and community standards that prohibit hate speech and put in place systems and teams to review content that is reported to violate these standards", and for them to have a system in place to remove notifications following their internal rules and community guidelines, and national laws implementing Decision 2008/913/JHA. With the Code, companies therefore commit to the following points:

- Review most content that has received reports within 24 hours and remove or disable access if it contains hate speech
- Train their staff
- Create partnerships and training activities with civil society
- Produce counter-narratives to hate speech and implement educational programs
- Establish contact points and appoint national contact persons to facilitate cooperation with relevant national authorities on the subject
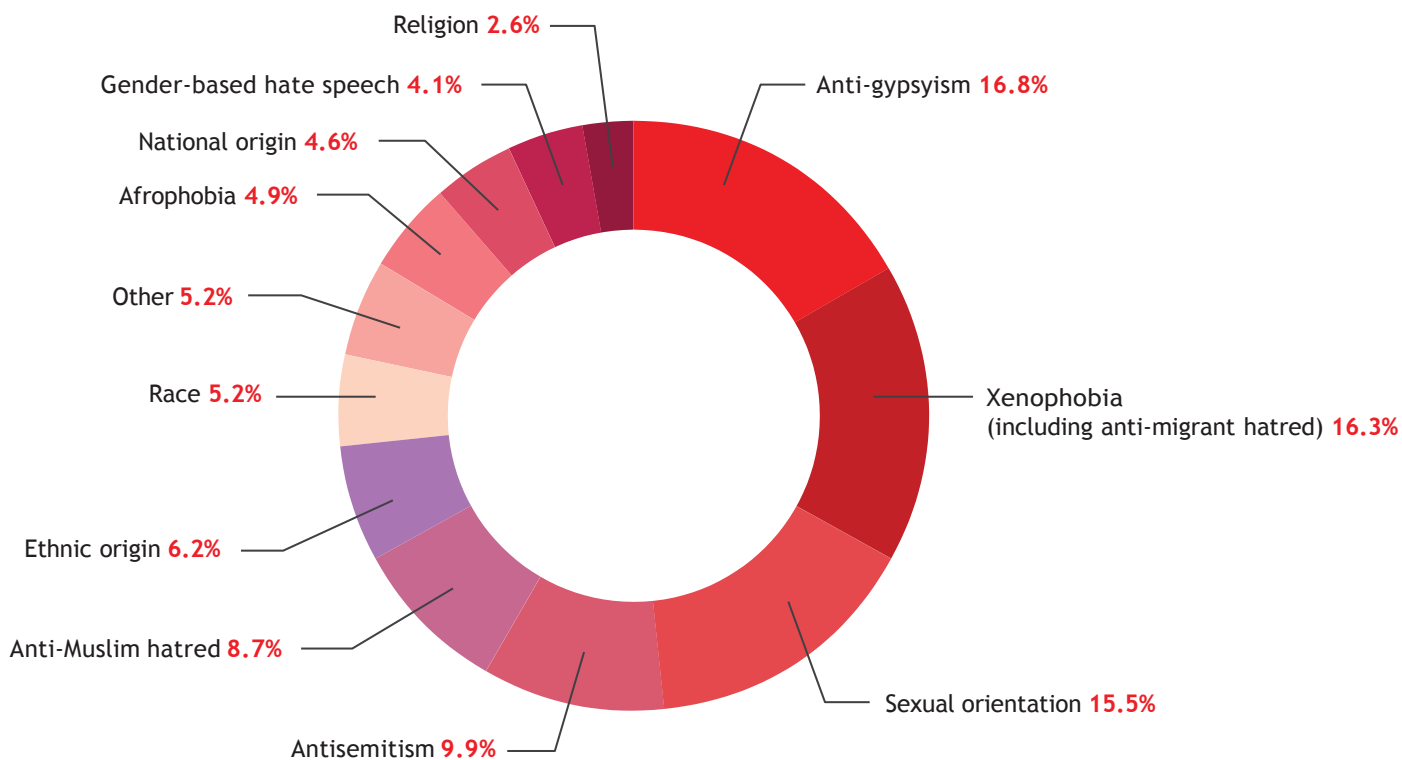- Promote transparency to users and the general public.

The final point that companies undertake is to individually define hate speech according to their corporate policy. In this case, for example, Facebook, then Meta, defines it as "a direct attack on people based on what we consider to be protected characteristics like race, ethnicity, national origin, religious affiliation, sexual orientation, caste, gender, sexual identity, and serious illness or disability." It then pledges to protect immigrant status and adds that it considers hate speech violent and dehumanizing, a statement of exclusion or segregation. But in the last part of its hate speech legislation, it explains that: "we also prohibit the use of insults aimed at attacking people on the basis of their protected characteristics. However, we recognize that people in some cases share content that incites hatred of which they are not the author or that contains insults for the purpose of condemnation or awareness-raising. In some cases, speech, including insults, that might otherwise violate our standards may be used in a self-referential manner or to reinforce a cause. Our standards are designed to leave room for these types of speech, but we ask people to clarify their intentions. When the intention is not clear, we may remove the content' (Meta Transparency Centre).

Another social media platform to consider is TikTok. In its guidelines, the company defines hate speech similarly to Meta but considers hate speech and insults as synonymous, defining them as "derogatory terms that aim to denigrate groups or individuals based on any of the protected attributes listed above" (refers to the categories under which hate speech falls) (TikTok Community Guidelines).

Twitter, in its 2023 policy, defines that one cannot directly attack a person for characteristics of their identity. It emphasizes that freedom of expression is a human right and that everyone has the right to use it, to foster public conversation, and thus accepta 'wide range of perspectives'. It does, however, take action (without specifying what kind) for hate speech, insults, dehumanization, hateful images, but emphasizes the fact that while some content may seem hateful, "it may not be hateful when viewed within the context of a broader conversation", so sometimes it may "not be clear whether the context is to abuse an individual based on their protected status or whether it is part of a consensual conversation" (X Help Centre, February 2023).

Since 2016, IT companies adhering to the Code of Conduct have progressively equipped themselves with automatic systems to detect and remove hate content, and in 2019, for example, Facebook's data show that, of all content removed, 65 % was done bythe algorithm. The problem to date is that, unfortunately, the algorithms are only able to recognize content with explicit hate speech (based on gender, race, language, religion, etc.), but often not those that contain strong allusions but do not contain words explicitly referring to the topics contained in the definition of hate speech (Council of European Union 2019). According to the latest annual reports on measures, both taken by IT companies on their work and monitored by the European Commission at the end of 2022, on average of all reported hate content only 63.6 % was removed, the rest remained online. YouTube has the highest content removal rate with 91 %, followed by Facebook (69.1 %), TikTok (60.2 %), Instagram (58.4 %), and finally Twitter with 45 % (EU Commission 2022).

The most frequently reported instances of hate speech in Europe in 2022 are anti-Gypsyism (16.8%), xenophobia particularly against migrants (16.3%), and sexual orientation (15.5%). Just below the 10% threshold, we find anti-Semitic and Islamophobic content, lower down is hate speech on ethnic origins (6.2%), race, Afrophobia, national origins, and sexism.

Religion **2.6%**

Gender-based hate speech **4.1%**

National origin **4.6%**

Afrophobia **4.9%**

Other **5.2%**

Race **5.2%**

Ethnic origin **6.2%**

Anti-Muslim hatred **8.7%**

Antisemitism **9.9%**

Anti-gypsyism **16.8%**

Xenophobia
(including anti-migrant hatred) **16.3%**

Sexual orientation **15.5%**

See EU Commission '7th evaluation of the Code of Conduct, November 2022''[1].

Council of Europe - Recommendation CM/Rec(2022)16[1]

The last Recommendation of the Council of Europe on hate speech was adopted in Turin (Italy) in May 2022. The CoE's Secretary General, Miss Marija Pejčinović Burić, stated: "hate speech is on the rise in Europe, especially online, where it often takes the form of racism, anti-Semitism or incitement to violence. European governments must join forces to tackle this complex threat to our societies by taking effective and proportionate measures". The document urges Member States to adopt different types of legislative measures depending on the severity of hate speech cases. Indeed, the Recommendation states that hate speech covers a wide range of expressions with different levels of seriousness and proposes that Member States adopt measures based on the seriousness of the offense. Thus, it is argued that a distinction should be made between incitement to hatred that reaches levels of seriousness prohibited by the Criminal Code and that can take various forms, from incitement to commit a crime against humanity (e.g. genocide) and incitement of the public to commit discriminatory acts, to threats, to insults of all kinds of what is called hate speech, to the dissemination of material. Other less serious types of speech should be subject to civil and administrative law, while it argues that others can be countered not by legislation but by educational and awareness-raising measures. The Recommendation then identifies points to refer to in order to calibrate the seriousness of a speech, such as the content, the social and political context in which it occurs, the

_____

1 Despite the recent policies to control the spread of hate speech on social media, only about 30 per cent was removed, more than 60 per cent remained in circulation. Which leads one to guess that a lot of hate speech was not even reported. The first chapter also pointed out that Meta and Twitter's main servers are located in the United States and the companies are owned by private US citizens. In order to comply with European directives in recent years, companies have trained artificial intelligence to automatically detect hate speech and delete it. The problem arises when the terms used to hatred are not so explicit and are not defined by the individual words, but by the meaning of the sentence and the context to which they refer to.

intention, the public role or status of the speaker, the possible harmful effects that may result, the size of the audience attending the speech and the characteristics of the target groups. Another aspect highlighted in the Recommendation is that States should create laws that "require Internet intermediaries operating in their jurisdiction to respect human rights, including hate speech legislation, and to apply human rights due diligence principles in all their operations and policies, and to take measures in line with existing frameworks and procedures to combat hate speech". To this end, social media monopoly companies should take several measures such as the rapid processing of data and respective reporting of hate speech to the authorities; then, increased cooperation with local authorities to provide evidence and assessments, if there are cases that need to be analysed more closely and report such speech. The Recommendation states that the media should have the role of strengthening social cohesion, peaceful dialogue, promoting multiculturalism instead of spreading hate speech or discrimination.

HATE SPEECH

This section will analyze the concept of hate speech. There are many definitions of hate speech. In general, hate speech can take many forms and it can be spread online and offline. It can be conveyed by various forms of expression, including cartoons, memes, objects, images, symbols, gestures, and fake news that give a distorted view of reality (United Nations). In addition, the context in which certain messages or content are produced is relevant.

The Council of Europe defines hate speech as "all kinds of expressions that incite, promote, spread or justify violence, hatred or discrimination against a person or a group of persons, or that denigrate them, because of their real or attributed personal characteristics or status such as race, colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation". The Council of Europe emphasizes that hate speech, by undermining the freedom of individuals, also entails an attack on democracy.

Another definition commonly referred to is that of the European Commission against Racism and Intolerance (ECRI) of the Council of Europe, which also identifies hate speech as "the instigation, promotion or incitement to denigrate, hate or defame a person or group of persons, or the subjection of a person or group of persons to harassment, harassment, insults, negative stereotypes, stigmatisation of or threats against such a person or group, and includes the justification of these various forms of expression based on a variety of grounds, such as 'race', colour, language, religion or belief, nationality or national or ethnic origin, as well as ancestry, age, disability, sex, gender identity, sexual orientation, and any other personal characteristic or situation." In recent years, the Council of Europe has devoted much attention to the problem and to finding ways to counter it, proposing not only appropriate actions to legally balance freedom of expression with other fundamental rights, but also many non-formal educational approaches to human rights and the recognition of the socio-psychological impact that hate speech has on individuals.

The definition to which the Hate Trackers Beyond Borders project mainly refers is that of the United Nations Hate Speech Strategy and Action Plan, which defines the term as "any type of oral, written or behavioural communication that attacks or uses pejorative or discriminatory language about a person or group based on who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor". This is often rooted in, and generates intolerance and hatred and, in certain contexts, can be demeaning and divisive" (UN Strategy and Plan of Action on Hate Speech, 2019). However, it must be emphasized that this is not a legal definition. As far as international law is concerned, there is no universal definition, but there are others that emphasize different points. The UN Strategy and Plan of Action defines hate speech as a form of discriminatory or pejorative expression of the characteristics described above and defined as identity factors, thus undermining an individual's identity, and emphasises the fact that current international law only prohibits incitement to discrimination, hostility and violence. The issue in focusing only on incitement is that, when derogatory language does not take the shape of incitement, States are allowed not to punish it (UN, 2019). Regarding the severity of penalties and the definition of the type of offence, there are legal differences. Each country has its legal code of conduct, which means that each country has different legislation and regulations that are then applied differently in their own country. Therefore, these differentiations lead to potential debates and gaps among academics dealing with the specific topic. The issue and the need to find a common and shared solution that has been raised in the last forty years is the rise of social media, of cyberspace as a parallel and unregulated virtual space that is deeply embedded in society and whose effects are widely visible in the offline sphere.

The Pyramid of Hate

To define hate speech, the Council of Europe (CoE) started with a pyramid analysis of what leads to this kind of language. The base of the pyramid is stereotypes. These are defined as shared beliefs or preconceptions about certain groups, which can be negative, positive, or neutral. Psychological studies argue that the use of stereotypes is not necessarily negative because it is an important way to simplify reality. After all, they allow our brain to reduce the amount of information we need to process when we meet someone or when we come across situations or environments (Mcleod, 2023). Sometimes they can be useful because they allow us to respond quickly to situations that are in front of us, they help to define what is around people, to define and identify with certain ethnic or social groups. However, they can also have negative meanings as they lead to ignoring differences between individuals, lead to untrue assumptions, and even justify violent actions towards a different ethnicity, sexual gender, or religion. In fact, they become harmful when they are used as a pretext to justify negative behavior or actions towards other groups.

On the second level of the pyramid, there are prejudices. A prejudice is nothing more than the application of a stereotype, or rather a stereotype that at its core already contains a premature evaluation or judgment about a person or social group. Unlike the stereotype, prejudice is a thought or behavior that is explicit or implicit, but that people are aware of having; it is a thought that is constructed from personal experiences, from those around

us, from the speeches we listen to, from readings, and from personal interests (Dovidio & al., 2019).

On the third rung, moving upward, we then find discrimination, which is considered the enactment of prejudice, and unfavourable treatment, and contains within it only negative elements. Social studies argues that discrimination stems from deeply ingrained and destructive generalizations of a certain group, such as discriminatory or unrealistic beliefs against a certain group of people that contain within them beliefs of intellectual and moral superiority of the group they belong to (Keita 2016).

There are many examples of this, the most common and well known being racial discrimination perpetrated for centuries and, speaking of modern history, racial, social, intellectual superiority that has been used as a tool to justify exploitation, wars, violence all over the world. Discrimination is therefore a behaviour that does not accept and respect certain groups, violating their human rights.

At the top of the pyramid, we find hate crimes. The Council of Europe describes hate crimes as "unlawful acts against a group or individual based on a bias against their perceived identity". They can include crimes against property, such as vandalism, and/or against people, such as bullying, harassment, physical violence, and murder. Genocides are examples of large-scale hate crimes.

The main consequences of a hate crime are post-traumatic stress, high levels of depression, anger to the point of committing crimes, fear, disintegration and deterioration of interpersonal and social relationships, and even isolation from society. In society at large, it has been noted that hate crimes are highly and more likely to create retaliation, riots, and revenge of equal or greater severity because, unlike other crimes, hate crimes take on a strong symbolic value. Therefore, they often take the form of large-scale urban riots (e.g., the Stonewall riots in New York City in June 1969). If hate crime is a criminal act, hate speech is a negative expression that may result or be considered by some as a crime to be punished criminally.

Findings already gathered from previous research while conducting the first #HateTrackers project have shown that if hate speech goes unchallenged, it pushes human rights violations further: negative stereotypes are spread throughout society, groups become increasingly marginalized and isolated, conflicts and divisions grow, and abuses or threats increase when new boundaries are tested. In the worst cases, mere "expression" begins to translate into physical abuse. Hate speech can lead to hate crimes, involving human rights related to personal security. The point the project intends to make about the danger of hate speech is that "not all hate speech can result in hate crimes, but hate crimes, including genocide, are always accompanied by hate speech." Explicative of this point is the Nazi propaganda perpetrated by Goebbels and Hitler during World War II and carried out through the polarization of the media apparatus in the service of propaganda. Another clear example the project emphasizes is the Rwandan genocide in 1994, in which the instrumentalization of radio played a crucial role in spreading hatred and inciting and encouraging people to participate in the violence that killed more than a million people in a few months.

The effects of hate speech

Several scholars have focused on the effects of hate speech. In a study by Cowan and Hodge (1996), for example, it was shown that derogatory speech delivered in public deprives/disturbs the moral integrity, personality, and identity of the target toward whom it is directed. Other authors have highlighted other effects, such as feelings of fear, disempowerment, exclusion, dehumanization, silencing, as well as feelings of anger and the exacerbation of discriminatory biases and the transmission of these to future generations (Gelber and McNamara, 2016). In this research of 212 college students selected based on ethnicity, gender, education, and sexual orientation, hate speech in public was considered more serious than hate speech delivered in private. The main problem was the performance of this task in public, thus arousing the desire to subordinate a group's identity, based on the idea that the group or person targeted for its identity factors is inferior to another group. Another author, Gelber, identifies public hate speech as a discriminatory act "that acts on its targets in constitutive and random ways to achieve the denial of equal opportunities and rights" (Gelber, 2021). For the author, violent speech must be regulated by considering the power of the speaker and the consent he or she can gain from the audience. Silence and the absence of criticism are also symbols of tacit consent. Targeting and humiliating people based on identity factors during a public debate can undermine participation and equal opportunity.

On the other hand, the danger of private hate speech should not be underestimated either, because it was shown that when the victim did not respond or ignored the verbal attack, this was seen as a condition of strong vulnerability and inability to defend themselves and react, and therefore gave the perception that the hater had more power than the victim. Even in the private sphere, therefore, conditions of subordination to the one being targeted occurred.

One of the most devastating effects of hate speech is the dehumanization of the victim. Other academics have shown that discrimination arising from constant exposure to hate speech can occur on several levels. One who is bombarded by fake news that portrays a social group in a certain way may be discriminatory towards the latter not only on an individual/personal level where he/she will simply keep his/her distance or distance those who belong to this group, or in a more serious form by using offensive language due to envy, anger or other types of emotions, but may also act by advocating harsh, exclusive contrary policies aimed at that group. On the side of the victim of hateful messages, studies have shown that frequent exposure to violence or violent behavior leads the subject to perceive acts of violence as less severe, less serious, which can lead to apathy towards violent events/facts, even to the point of subconsciously considering violence as the norm (Bilewicz & Soral, 2020).

To conclude this section, it is also worth mentioning the Psychological First Aid Toolkit aimed at all those who suffer or are victims of hate speech. The toolkit was developed by professionals from Hope For Children, including activists from the Hate Trackers project. It presents its effects and causes, the challenge in counteracting, recognising, and identifying it, and gives some tips on how to react personally and process exposure to hate.

The risks that online activists can face in pursuing their actions

In the previous section, we talked about the effects that frequent exposure to hate speech can have on victims, witnesses and haters alike. For a more in-depth look at its effects, Hope For Children prepared the [Psychological First Aid Toolkit (PFA)](#) exploring the consequences of the frequent exposure to hate speech on mental health. One of the main effects is that such derogatory speech and language, if not promptly interrupted, leads to de-humanization and a lack of empathy towards the targeted group. The major consequence is the creation of a social norm that legitimizes and accepts this kind of speech, and people will be more inclined to use derogatory language, while on the other hand it will lead to mental breakdown of the victims. The PFA also suggests how to react to hate speech, by taking action, reporting, seeking external support, but also by taking care of oneself, and finally emphasises the adaptation of the concept of Nonviolent Communication.

Nonviolent Communication

The concept of Nonviolent Communication was first defined by Marshall Rosenberg, who founded the Centre for Nonviolent Communication in the early 1960s. The key factors of his model are empathy, sincerity, communication, listening, sharing feelings and personal needs without judging. He believed that "human beings will only be ableto communicate effectively in a non-violent manner when they are able to express how they are and how they would like to be, pointing out specific behaviour to the other, without confusing facts with opinions" (Costa Pinto & Nascimento Cunha 2023).

The model metaphorically divides violent language (also called jackal language) from non-violent language (giraffe language). Violent language is characterised by a lack of empathy, it takes the form of judgement, criticism, evaluation of what is good or bad, right or wrong, based on self-perceptions that do not take into account the needs of others.

Non-violent communication, on the other hand, is based on empathy, connection, expressing one's needs and listening to others. It is about giving and receiving at the same time. The first step of the model consists of observing without judging, but understanding one's own feelings and those of one's interlocutor. The second step consists of expressing one's personal emotions openly. The third step is to express the needs arising from one's feelings. The fourth and last step consists of communicating our requests being careful not to express a need; spontaneity is a crucial aspect of this part of communication.

Risk assessment of online activism

In the digital environment, there are certain risks that online activists may face. Among those analysed by academics and journalists, the most recognised are:

1. **Digital gap**. Not everyone has the same tools to access online content and the possibility to spread his/her/their own ideas as much as who can afford to spend a higher capital on social media. The latter refers to the social media sponsoring/ social media campaigns: whoever can afford to spend money, can pay a sponsorship to one of the social media companies, thereby the algorithm will make the content visible to a wider number of users.

2. **Unfair democratic participation and instrumentalisation of social media for political purposes.** Thanks to social media sponsorships and the policiesof technology companies that allow for broad freedom of expression, political parties and private users who can spend large resources will have greater visibility of their posts. For example, powerful and wealthy political parties or private users run political/social campaigns aimed at targeting, marginalizing, and discriminating against a specific group of people. But if the targeted group or those who support it do not have the same number of economic resources, their response will be heard/seen by fewer users, resulting in unequal participation in the democratic debate.

3. **Slacktivism.** This term is used to define the type of lazy activism that does not require political engagement/self-commitment and exposure but may be limited to sharing or retweeting posts/hashtags/stories without carrying out real political/ social action. According to Forbes, this behavior lacks real political action and can be detrimental to both the cause and the 'slacker' (Travers, 2023). The journalist further argues that "when online activism is practiced without intention or critical awareness, it can promote slacktivism, which is often criticized for requiring onlya minimal display of support for a cause, as well as lacking the intention to engagein enacting meaningful change". In terms of psychological repercussions, being frequently engaged in slacktivism can lead to:

   - Emotional exhaustion or emotional emptying, which stems from the fact that this form of activism is only seen as a form of show and appearance rather than real action to bring about change, which can end up disconnecting from one's real feelings.
   - Hypocrisy between espousing values and real actions can lead to feelings of guilt.
   - The creation of an unrealistic sense of self-fulfillment. Slacktivism only gives the illusion but does not produce any significant change.
   - Stress and anxiety, such as attempting to curate form/profile on social media without delving into actions, can lead to the need to seek social approval and receive positive interactions before pursuing real change.

4. **Polarisation.** The Cambridge Dictionary defines polarisation as the act of dividing something, people, or opinions into two completely opposite groups.It is a multifaceted social phenomenon that can take two main forms: ideological and affective. The first is the process that leads to the radicalisation of opinions on certain topics, mostly political and social, resulting in the hardening of ideas and ideological distance in society. While the second refers specifically to the construction of social identities and socio-emotional distance between groups, which leads to phenomena such as distrust, growing aversion, mutual hostility and the 'us versus them' paradigm.

Social media and their backbone, the algorithms, are greatly facilitating this process, leading to the exacerbation of social tensions, establishing and fostering a hostile communication terrain between parties, and sometimes even making it very difficult to reach out to the other side and exchange personal and differing opinions, thus eroding the core of democratic values.

In addition to the four issues mentioned above, there are others, also of popular origin, to which attention should be paid, not only during online activism, but also when planning actions.

1. **Education and activism are not the same.** Education is the basis of activism, it provides the tools, the knowledge, it is responsible for developing critical thinking. On the other hand, activism is defined by the Cambridge Dictionary as "the use of overt direct action to achieve a result, usually political or social". Activism is direct action, according to the Compass Handbook "when applied to human rights, it means defending human rights wherever they are threatened or violated, and at whatever level. Human rights activism is therefore about reacting to injustice, abusive treatment, violence or discrimination and seeking to correct them. It is about being prepared to assist and show solidarity with the struggles of others, to fight to ensure that they are treated with respect and dignity, and to help facilitate the transition to a more humane, egalitarian and rights-respecting society' (Compass Handbook, p. 360).

2. **Quality activism requires planning.** Having sessions/meetings in which you set your goals, the means/actions to achieve them, a careful analysis of the material and everything you need to carry out your actions. But above all, you have to draw up a detailed analysis of the social, geographical, and political context in which you are planning your actions, and adapt them according to it, trying to achieve themain objectives.

3. **Be aware of risks.** Not being aware of the possible risks and obstacles you may encounter is a risk in itself, especially when taking action online.

4. **There is a close connection between the advent of social media and the rise of hate speech.** As already mentioned, social media were not created with the main and only intention of connecting people, but to sell. As far as the dissemination of ideas is concerned, algorithms tend to lock social media users into media bubbles from which it is very much, but not impossible, to get out. One of the major

consequences, as already mentioned, is an increasing polarisation of ideas.

5. **The real effectiveness of an online campaign.** Since the media tend to close themselves in a media bubble, also called 'echo-chambers', it is also quite complicated to get out of it, to be more effective and reach more users. Indeed, today's algorithm can turn their feeds into echo-chambers of divisive content and news of varying reputations that support their perspectives' (Oremus et al. 2021). This is why it is very useful to study a bit or take examples from other successful online campaigns (e.g. study the hashtags or network they created before acting online).

6. **Nobody knows whether the functioning of social media is still in the hands of a few people or whether artificial intelligence is already out of the control of mankind.** In the first case, the issue is that the key values of democracy are not respected. The second is equally problematic because no one can predict who will see your campaign/post/action online.

7. **Many argue that their structure resembles or at least has many similarities to the Nazi propaganda of Hitler and Goebbels during the Second World War.** Misinformation and fake news, posts/tweets that arouse strong and mixed emotions circulate more on social media and reach more users. The posts themselves, it has been studied, are the best breeding ground for hate speech.

8. **Shit storm over your post.** Be aware that your post may generate support, but also hostility. Social media is mostly an unregulated cyber space, where those who spread hateful/discriminatory content and/or hateful comments are hardly ever punished.

9. **Be aware that what happens in the online world has repercussions in the offline sphere.** The two spheres are closely interconnected and linked to each other. A relevant and classic example of this point is the protests that took place after the murder of George Floyd in the United States: the online campaign turned into social riots and protests across the country. But in everyday life, it can happen that at school a shit storm or an act of cyberbullying carried out against a student can have physical and social repercussions in offline reality.

10. **Anonymity and the use of false identities.** On social media, interactions are mediated by a screen. Interactions that take place on social media are not direct, but users are protected by a screen and can therefore use false identities or even steal someone's identity, being careful to report it if necessary.

Public hate speech and the role of social media

Before introducing and discussing hate speech on social media, the structure, causes and possible consequences of offline hate speech will be explained. More specifically, how offline public hate speech can be dangerous and unregulated by stakeholders will be presented. Gelber argues that hate speech that acts on its targets in a constitutive and casual manner will lead to the denial of equal opportunities and rights in democratic participation (Gelber 202). According to the author, the focal point in determining the violent speech that needs to be regulated lies in displaying the speaker's power and ability to do harm. In some cases, publicly belittling targets of people based on their identity factors can jeopardize equal participation and equal opportunity in the dialogue and thus the democratic instance, resulting in placing them in an inferior position in which their opinions will be of less value than those of the aggressor.

The speaker doesn't need to be a famous public figure, and how he or she can gain power can vary. This, for example, may come not only from the consent of the audience to whom he or she is speaking, but also from silence and the absence of criticism. Other elements may be social status or the speaker's role in a specific context. Another element may be the speaker's actions and informal rules established. The same author thus agrees that "whena person delivers a speech that reinforces and perpetuates existing systemic discrimination against a marginalized minority, that person has the ability to enforce his or her speech act as an act of oppression by virtue of the fact that it took place in a society that is imbued with such systemic discrimination." The issue is the permissibility of discrimination within the dialogue, and thus its informal normalization. If speech that is deemed violent or derogatory is not challenged/moderated or criticized, but passively accepted or ignored, it ends up establishing a new boundary of what is allowed to be said within public discourse, which will then create an informal norm in what is allowed or not allowed to be said. If what is allowed to be said creates a relationship of subordination on the part of one group over another, this is where what is allowed to be said is normalized and thus discrimination is legitimized, and this is where normative, social and educational actions must be taken.

But the above is nothing new. Just think of many historical facts including the period of the first colonizers on the African continent and the systematic racism that was carried out by politicians, scientists including Darwin himself when he spoke of the evolution of the human race and where he contrasted the inhabitants of the continent with "Western man," religious missionaries who set themselves the task of civilizing peoples but did nothing more than destroy millennia-old cultures on the basis of a conception of ethnic superiority. This not only normalized the thesis that there was a superior and an inferior race or ethnicity, but also produced devastating effects in the translation of racist behavior into laws (see racial segregation in the United States).

Language has power, and the power derived from it can be of gigantic dimensions. Accepting and establishing what can be said as it has been said sets limits; therefore, it will allow other speakers to conduct speeches of that certain type. Many speeches of that particular type, if delivered frequently and if used or confirmed by personalities with different roles (e.g., politicians, scientists, anthropologists, economists, etc.), will be incorporated into culture

and traditions. They will then be passed on to future generations who will consider it normal to conceive of one group as subordinate or inferior to another. Thereafter, not only hate speeches will be legitimized, but also large-scale actions will be publicly supported, up to the most extreme case of convincing a large segment of the population to commit genocide by justifying it as a just cause. This is where the heart of the problem of the systematization of hate speech, as the author identifies it, lies, and it is in these cases that, in her view, regulations or norms need to be put in place to prevent the spread of such speech. In the case of public hate speech, therefore, another key feature is that it occurs within a social and institutional context in which systemic discrimination against a social group is perpetrated and which helps to introduce or reinforce social inequality in the circumstance in which it occurs and legitimizes it. However, the author points out that societies are not static.

Studying the social media, some scholars have argued that "the recent proliferation of discriminatory ideologies and far-right political groups and parties has been accompanied by the rise of social media" (Donzelli, 2021). The rise of social media has been accompanied by the proliferation of racist, discriminatory, sexist, xenophobic, misogynist, and homophobic ideologies perpetrated mainly by far-right political parties and groups. A crucial factor in the dissemination of hate messages is the wide range of choices with which online content can be created. For example, on social media it is possible to convey hate messages through memes, emoji, reels, gifs, or exploit anonymity through fake profiles. Another tool that can be used to produce and spread discrimination are filters used to edit photos. Another tool used nowadays, especially by political parties, is also the use of artificial intelligence to create fake photos. It is equally possible and quite easy to steal someone else's digital identity and spread hatred under the guise of that person, without putting oneself in the position of spreading hate language. Another important reason is the de-personification of the victim and the lack of empathy caused by physical or real distance, and the hater's sense of protection behind a screen. The dynamics that are created on social media are different from those offline, the structures and basis of hate speech are different. Earlier it was explained how hate speech can turn into an informal norm, taking into consideration the speaker's formal or informal authority and power granted by the audience based on interactions, consensus and dissent, and criticism. In the digital sphere this top-down approach does not exist, rather the discourse gains media relevance based on interactions to the post and/or commentary through likes, comments, and reactions (referring to the structure of Facebook where instead of a like one can add various types of reactions, including emoticons representing disgust and anger). In the online world, it does not matter so much that the post receives many likes or approvals, but only the number of interactions that post, or comment, produces, whether positive or negative. This is what interests the artificial intelligence that governs the structure of any social media, the algorithm. For the algorithm, the more interactions a post has, the more it will gain relevance and it will be viewed by more and more users. The algorithm was created to regulate social media, so it is necessary to ask what the purpose of social media is. The algorithm is a set of rules that rank and determine, in order not chronologically but in order of importance, the content visible on social platforms. The algorithm evaluates each post, comment, story based on the interactions it receives, but not only that; using a smartphone, which works based on the touch screen mechanism, the algorithm is able to calculate the microseconds each user spends looking at that post/video/review. Through these mechanisms, added to the processing of personal data and

the way it is processed, and which by subscribing to any platform each user accepts, the algorithms can determine the personal feed (Oremus et al., 2021). This is how each social media outlet tries to organize the content visible to each user to show the content that, based on the data collected, it deems of greatest interest. The reason is simple, the social media wants users to spend so much time on it, the underlying reason: to monetize. The main purpose of social media platforms is to monetize, not to connect people, and the way they are successful in doing this is through advertisements/sponsorships that users, for various reasons, pay for, which is a mechanism by which posts become more visible based on how much is invested. The purpose of social is to sell (Brown 2021). The way: to show personalized advertisements based on the data we allow algorithms to collect every time we connect to Facebook (for example). In addition to the means by which Facebook's algorithm is able to figure out the type of advertising content each person is interested in, there are also the "favorites" option, i.e., the pages/posts/people we are most interested in viewing, and "in-feed options" that allow the user to express a personal level of liking for each post.

Briefly analyzing how Facebook's algorithm determines what content is relevant and what is not, the following are evaluated: who posts it (the profiles with which each user has the most interactions), the type of content (made based on predictions about what the greatest interests might be), then the interactions of each post and their relevance (Newberry 2023).

While interactions and the relevance each post gets from these are key to the circulation of a post on social media, the problem with hate speech or negative or discriminatory speech, as the data show, is that it has been seen to generate more interactions than positive or neutral speech. Psychological but also political studies have shown that the posts that generate more interactions are those that can generate strong, even mixed emotions, so users will be more likely to react, comment, share, copy the link to send it to other users on other platforms. Thus, without a study of how algorithms work, the same users who are againsta post who comment on it, repost it (etc.) will allow it to circulate more on the network and thus be seen by other users, finding among them who will be for it and who will be against it. If it is hate speech or if the comments under the post contain hate, it will be seen by more and more people creating for the victim(s), making the effects exponentially worse. What has been said here has also been reported by several media outlets such as the Washington post (2021), where it states that since 2018, "the algorithm has elevated posts that encourage interaction... this not only prioritizes friends/relatives and pages that people interact with the most, but also gives it to divisive content." Another point about how the algorithm is among those responsible for the circulation of hate on the web is, the article says: 'each user's feed reflects the interests expressed. For a subset of highly partisan users, today's algorithm can turn their feeds into echo chambers of divisive content and news of varying reputations that support their perspectives' (Oremus et al., 2021). Another feature that has been noted is that the more a user interacts, even if only by watching videos often or scrolling througha news page for a long time, the more they will have access to viral content than those who use it less.

The first to campaign through social media was Obama during the 2008 presidential election campaign. Obama's election campaign represented a turning point for the new idea of propaganda because it was the first major online political campaign. The president

discovered the potential and effectiveness of social media (particularly Facebook) in spreading messages, the key role of algorithms, and the possibility of using them not only for commercial purposes but also for social and political ones. Obama marked a turning point for the development and effectiveness of propaganda today. Propaganda is the dissemination or disclosure of information, facts, arguments, rumors, or lies that influence public opinion. It is a form of manipulation and guidance that aims to guide people's beliefs and opinions. The influential content that is spread and replayed daily on social media platforms has the fatal consequence of influencing or even brainwashing readers, leading to hate speech. Misinformation can increase people's exposure to risks and vulnerabilities. Similarly, online hate speech invites violence against a minority group and can contribute to psychological and social harm through harassment, defamation, and intimidation. Propaganda uses these tools of disinformation and misinformation to promote hate speeches that are circulated to social media platforms.

LEGAL NOTES AND THE CONTEXT OF CYPRUS

According to the latest report published by the European Commission against Racism and Intolerance (ECRI), public discourse in Cyprus has become increasingly racist and xenophobic, particularly during election campaigns. An important feature of Cyprus is the strong influence of the Orthodox Church in political and public debate. Many reports published by KISA, a Cyprus-based NGO, have highlighted numerous hate statements made by the archbishop on various topics, especially on homosexuality and migration.

As studies have identified, hate speech in Cyprus has three lowest common denominators: inter-communal, inter-alterity, and inter-gender.

Inter-communal hate speech is mainly due to the division of the island. According to Dilmaç et al. (2021), the unsolved conflict has brought growing discriminatory behaviors and intolerance towards the Turkish side, which has triggered increased hate speech also towards vulnerable groups such as migrants, asylum seekers, Muslims. An example reported by the local newspaper Kathimerini in 2017, is a statement of the Archbishop of Cyprus in 2017 that described refugees as a threat to the nation that were sent to Cyprus to alter its national and cultural identity and to shift the demographic structure of the Republic (Kathimerini, 25/12/2017). According to the authors, hate speech is mostly expressed through elements of "otherism" and nationalism. They state that public figures, including politicians, members of the government, church officials, and journalists, often employ hate speech in their statements, without any consequences whatsoever. the authors reported that mainstream media regularly make use of hate speech, especially against migrants and refugees, while the use of discriminatory speech in social media is uncontrollable.

Hate speech is reproduced through distorted interpretation of facts, utilising narratives of "us versus them". Reports argue that public statements of hate made by religious, political and public actors are a common phenomenon, especially during election periods, in an attempt to garner public support. According to Dilmaç and other authors, politicians and

public figures resort to hate speech that incites hate crimes, violations of human rights, deep social divisions and inequalities, the marginalization and exclusion of the weak and the vulnerable. They make use of oversimplifications and fake news to appeal to people's basic instincts, popular resentment, distrust, and discontent, fears, feelings ofa widespread sense of insecurity and national identity, and apprehensions about their and their children's future (Dilmaç et al., 2021).

As already pointed out, hate speech is used as a tool to dehumanize, isolate, and discredit political competitors, human rights activists, civil society organizations, and anyone who shares cultural characteristics with the discriminated group, undermining their democratic participation.

Another topic fomented by nationalist groups and media is the **anti-migrants hatred**, especially since Cyprus joined the EU in 2004. Since then, the alarmist rhetoric and fake news towards migrants, especially the ones coming from the African continent and Middle East countries, has rapidly grown. This led to an increasing number of cases of racially motivated violence, also fomented by the far-right political movement ELAM. Portrayals of migrants in the mass media aim to create fear, hostility, and xenophobia. Migrants and refugees are often used as scapegoats for social and economic struggles and to distance society from the real causes of problems. Racist attacks on migrants are often censored, and local media have been criticized for generalizing and stereotyping the image of "migrants," thus contributing to dehumanizing them in the public's awareness and worsening the violation of their human rights (KISA, 2017).

KISA has published many reports on hate speech cases, and the Anti-Discrimination Authority and the Cyprus Media Complaints Commission (CMCC) have stated that "hate speech is dangerous because its effects and impact on individuals, society, and democracy cannot be underestimated or ignored: Human dignity, human rights, the claim of equal opportunity and equal participation, the idea of open and peaceful coexistence, and the democratic values of modern civilization are at stake." The Friedrich-Ebert-Stiftung has also reported on the inter-gender nexus of hate speech in Cyprus. LGBTQIA+ and gender-based discrimination are still widespread topics. Being portrayed as a conservative society rooted in religiousness and conservatism, the LGBTQIA+ community is considered a threat to moral and social values. LGBTQIA+ and gender discrimination are still widespread issues. Being portrayed as a conservative society rooted in religiosity and conservatism, the LGBTQIA+ community is seen as a threat to moral and social values. Gender-based violence stems from the male-dominated national struggles that determined social norms in Cyprus, and women's right to choose a different lifestyle was considered "out of place (Dilmaç et al.,2021). The phenomenon and issue of patriarchy is widespread across the Mediterranean region, and the Cyprus problem has dominated all aspects of Cypriot society and has contributed to making national patriarchies and traditional gender roles persistent, which leaves little to no space for women's movements to develop (Munt, 2020). According to the 2023 Global Gap Gender Report, compiled by the World Economic Forum, in a research that includes 146 countries (including all EU member states) gender equality has not been achieved anywhere. The closest to reach it are Iceland, Norway, Finland, New Zealand, Sweden, Germany, Nicaragua, Namibia, and Guatemala, having closed at least 80% of their gap, while the global average distance stands at 68.4%. The research analyzes four categories: Economic Participation and Opportunity, Educational

Attainment, Health and Survival, Political Empowerment. Cyprus is ranked 106th, with a gender gap index of 67.8% (where 100% would mean the achievement of full equality). Despite achieving almost full gender parity in education and health and survival, the data show that economic participation and opportunity is only 65.2 %, while political empowerment is only 10.9 % (Global Gap Gender Report, 2023, pp. 157-158).

To conclude, according to the bibliography, hate speech in Cyprus is considered a temporal issue rooted in the country's historical background (Dilmaç et al., 2021). It is related to its rich and complex history, political and social; to British colonization; to the struggle to build and consolidate a national identity; to relations and conflict with Turkey; and to the division of the island. Reports indicate that hate speech has a greater harmful reach than the official data suggest. As highlighted in the next section, there is no standard local legislation to counter and regulate it. Hate speech, discriminatory acts and hate crimes often go unreported and insufficient measures have been taken to regulate them. Hate speech is entrenched at the level of politicians, in the narrative structure of the media. It is spread on social media as part of a weak solution and negotiated between national and international institutions and high-tech industries.

The national legislation on discrimination and hate speech

This section provides an analysis of the national legal system regarding hate speech and discrimination. It delves into the Cypriot legal framework to provide a better understanding of the legal response to hate speech.

- **Law 134 (I)/2011 "The Combatting Certain Forms and Expressions of Racism and Xenophobia by means of Criminal Law".** The law criminalizes any person who deliberately transmits in public and publicly incites violence or hatred against a group of people or a member of a group, which is determined on the basis of race, color, religion, genealogical origin, national or ethnic origin in such a way to cause public disorder, or that has a threatening, abusive or offensive character. Xenophobic and racist reasons are considered as aggravating elements. Up to 5 years in prison and 10,000 euros fine are foreseen. The law only punishes conduct that causes or could cause public disturbance.

- **Law 84(I)/2003, Section 47(1)(b)** penalizes the public actions aimed at fomenting the enmity between the communities or religious groups on account of race, religion, color, or gender. Imprisonment up to 5 years, fine up to 20,000 euros.

- **Law 87 (I)/2015 amendment of the Penal Code, Section 99A** defines the "Incitement to violence or hatred base on sexual orientation or gender identity" as "(1) A person who intentionally, publicly and in a manner that is threatening or abusive or offensive in nature urges or incites, orally or through the press or with written texts or illustrations or in any other way, violence or hatred directed against a group persons or a member of a group of persons identified on the basis of their sexual orientation or gender identity, is guilty of an offense and, upon conviction, is liable to

imprisonment for a term not exceeding three (3) years or to a fine not exceeding five thousands of euros (€5,000) or both of these penalties.

- **Law 12/1967 and its amendment.** This act was meant to ratify the UN International Convention on the Elimination of All Forms of Racial Discrimination, adopted in 1965 and entered into force in 1969. The Convention defines racial discrimination as: "any distinction, exclusion, restriction or preference based on race, color, 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" (article 1). States ratifying the Convention must act to eradicate racial discrimination in every form through all legislative, political, educational instruments[2]. On its Amendment of 21 of February 1992, ratified by Cyprus on the 28 of September 1998, article 2 was amended, defining the term "Criminal offences" as "any person who publicly, whether orally or through the press or by written texts or illustrations or in any other way, intentionally incites acts or actions that can cause discrimination, hatred, or violence against persons or groups of persons , by reason alone of their racial or national origin or their religion, is guilty of a criminal offense and liable to imprisonment for a term not exceeding two years or to a fine not exceeding one thousand pounds or to both. (2) Any person who constitutes or participates in organizations, which pursue organized propaganda or activities of any kind tending to racial discrimination, is guilty of a criminal offense and is subject to the penalties provided for in subsection (I). (3) Any person who publicly, whether orally or through the press or through written texts or illustrations or in any other way, expresses ideas offensive to a person or group of persons because of their racial or national origin or their religion, is guilty of a criminal offense and shall be liable to imprisonment for a term not exceeding one year or to a fine not exceeding five hundred pounds or to both. (4) Any person who professionally supplies goods or offers services and refuses to provide them to another solely because of his racial or national origin or his religion or makes the supply subject to a condition related to the person's racial or national origin or religion is guilty of a criminal offense and is liable to imprisonment for a term not exceeding one year or to a fine not exceeding five hundred pounds or to both".

- **Constitution of the Republic of Cyprus:** Art. 19: Every person has the right to freedom of speech and expression in any form. This right includes freedom to hold opinions and receive and impart information and ideas without interference by any public authority and regardless of frontiers.   The exercise of the rights provided in paragraphs I and 2 of this Article may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary only in the interests of the security of the Republic or the constitutional order or the public safety or the public order or the public health or the public morals or for the protection of the reputation or rights of others or for preventing the disclosure of information received in confidence or for maintaining the authority and impartiality of the judiciary. Seizure of newspapers or other printed matter is not allowed without the written permission of the Attorney-General of the Republic, which must be confirmed by the decision of a competent court within a period not exceeding seventy-two hours, failing which the seizure shall be lifted. Nothing in

2

this Article contained shall prevent the Republic from requiring the licensing of sound and vision broadcasting or cinema enterprises. Art. 28 par.2: Every person shall enjoy all the rights and liberties provided for in this Constitution without any direct or indirect discrimination against any person on the ground of his community, race, religion, language, sex, political or other convictions, national or social descent, birth, colour, wealth, social class, or on any ground whatsoever, unless there is express provision to the contrary in this Constitution.

Law Enforcement Agencies

- Police Office for Combating Discrimination (OCD) oversees monitoring the reports and complaints of cases of discrimination submitted to the police and investigates them.
- Cyprus Media Complaints Commission (CMCC) is an independent body that examines and oversees the code of conduct of journalists on hate speech and discrimination.

Same thing goes for "The Cyprus Radio Television Authority (CRTA)" which is responsiblefor monitoring radio stations and electronic media. Though it does not include social media.

LOCAL SERVICES IN CYPRUS TO SUPPORT VICTIMS OF HATE SPEECH

Hate speech has detrimental effects on people's mental health and well-being. Frequent exposure to derogatory language can cause very dangerous effects. It can also lead people to feel apathy toward it and accept it as a social norm. This can lead to a lack of empathy toward targets, even to the point of dehumanizing them.

People exposed to hate speech may develop post-traumatic stress disorder, lack of self-confidence, low self-esteem, anxiety, depression, or worse.

As part of the Hate Trackers Beyond Borders project, Hope For Children's specialized staff developed a Psychological First Aid Toolkit aimed at victims and people exposed to online and offline hate speech. It provides self-care tips and practical exercises for dealing with stressful situations.

Extensive desk research and national mapping were conducted to identify local services in Cyprus where victims can seek psychological help and support.

The list is available in English.

CYberSafety Youth Panel	The 1480 Complaints Hotline offers a direct, easily accessible and accountable point of contact for users to report illegal content or actions, including racist and xenophobic material. The 1480 Helpline services are aimed at children, teenagers, parents, teachers and other professionals, providing support and advice on issues related to: • Cyberbullying • Online Racism and Xenophobia • Problems in social networks (e.g., fake profiles, account hacking, inappropriate and harmful content) • Safe and responsible use of internet. ☎ **Helpline: 1480** ✉ **1480helpline@cyearn.pi.ac.cy**
Cyprus Samaritans	Free service to provide confidential emotional support to adults who are feeling distressed, lonely or unable to cope. ☎ **80007773** ✉ **confidential@cyprussamaritans.org**

"Hope for Children" CRC Policy Center	The 1466 Helpline is available to all children and families in Cyprus free of charge, 24/7 and provides psychological, social and legal support, advice and guidance on issues related to child protection. The 116 111 European Helpline for Children and Adolescents is operated by HFC and the Association for the Prevention and Handling of Violence in the Family (SPAVO) and provides the opportunity to youth to talk about their concerns and worries. ☎ **Helpline: 1466** ☎ **European Helpline for Children and Adolescents: 116 111**
Youth Organization of Cyprus (ONEK)	Free service to provide emotional support to young people experiencing negative feelings and challenging situations. ☎ **Helpline: 1410**

Nicosia

The Centre for Therapy, Training, and Research (KESY) Student Counselling Services (University of Nicosia)	• Individual and Group Psychotherapy • Family Therapy • Therapy for Children and Adolescents • Parent Counselling for the Education of Children and Adolescents • Psychological Tests and Diagnostic Assessments • Training Seminars for Parents and Professionals ☎ **22 795100**
Nicosia Municipality Counselling Centre	• Personal meetings with specialist staff or referral to specialist services if necessary. • Organization of group experiential workshops to develop social skills to better address the needs of everyday life, exchange of experiences, development of self-confidence. • Organizing lectures to inform and encourage prevention in the community. The topics of the lectures include topics such as: parent-child relationships, learning difficulties, old age, drugs, mental health, sexual health, diabetes, cancer, work relationships, etc. ☎ **22 797870**
Nicosia Multipurpose Municipal Centre	Support and guidance from professional psychologists, counselling psychologists and social workers at individual, group and community level ☎ **22 797850**

Counseling Station Ithaca (Nicosia)	Counseling and/or psychological therapy services are provided to individuals, couples and families who are experiencing any mental health problem that can be treated on an outpatient basis (does not require hospitalization). ☎ **22 521570**
Strovolos Municipality Counselling Centre	Provision of psychological support. ☎ **22 470374 / 22 511728**
Youth Information Centres Programme Proposal (Nicosia)	Provision of psychological support. ☎ **97 725204**
Child and Adolescent Mental Health Services (Archbishop Makarios III Hospital)	Diagnostic evaluation and therapeutic planning and intervention, depending on the individual case, in children and adolescents up to 17 years of age, as well as support and counselling for parents. ☎ **22 405085**
Center for Adolescent and Family Counseling "PERSEAS" (Nicosia)	Intervention with therapeutic and counselling approaches to treat addiction to illegal substances/alcohol/gambling/internet. ☎ **22 464292**
Therapeutic Community "Agia Skepi" (Nicosia)	Intervention with therapeutic and counselling approaches to treat addiction to illegal substances/alcohol/gambling/internet. ☎ **22 464292**

Limassol

Family Guidance Centre Metropolis of Limassol	• Direct intervention in crisis situations. • Strengthening the institution of the family. • Developing the skills of family members to enable them to fulfil their conjugal and parental role. • Prevent and address the problem of school leakage, which is most prevalent in the area where the Family Guidance Centre operates. • Providing support and counselling services to individuals and families. • Training adults in skills development to address individual and family problems. • Providing primary and secondary prevention services to address the phenomenon of school drop-outs. 📞 **25 864315 / 25 864330**
Youth Information Centers Programme Proposal	Provision of psychological support. 📞 **97 725204**
Municipality of Mesa Geitonia	Free psychological treatment and support for adults, children and families. 📞 **96 660076**
Center for Prevention and Counseling of Adolescents and Families "PROMITHEAS" - Department of Mental Health Services.	Provision of therapeutic programms for psychological rehabilitation. 📞 **25 305079 / 25 305110**

Larnaca

Larnaca Child and Adolescent Mental Health Services/ Famagusta	• Services to children and adolescents up to 17 years old with various psychosocial difficulties. • Counselling and guidance to parents and families. 📞 **24 813164 / 24 813174 / 24 813074 / 24 813182**
Aradippou Municipality	Free psychological support services. 📞 **99 951218**
Youth Program Proposal	Psychological support services. 📞 **97 725204**

Pafos

Neapolis University Counseling Center (S.K.E.P.S.Y.)	The team of the Counselling Centre for Research and Psychological Services (C.R.P.S.S.) of Neapolis Paphos University offers its services to the students of the University as well as to the wider community of Paphos (minors and adults). The services of the Centre are free of charge. ☎ **26 843425 / 96 417024**
Youth Information Centres Proposal	Psychological support. ☎ **97 725204**

CONCLUSION

Hate speech is a persistent problem in Cyprus and undermines fundamental humanrights. The report delved into the definition of hate speech, the different approaches to the phenomenon, and the legal framework in Cyprus, showing the gaps and challenges to be addressed.

According to desk research and previous studies, the Cypriot legal system in place is considered inadequate and selectively applied to hate incidents. The criminal justice system's response to hate crimes is not effective. Criminal law provisions against racist hate speech are not enforced and perpetrators (including public figures, government ministers, and church officials) face no consequences, which sends a message of impunity (Dilmaç et al., 2021). Hate speech incidents are either not identified and recorded properly or, even if properly recorded, they are not prosecuted most of the time, because they are not substantiated, according to the authorities (Dilmaç et al., 2021).

There is inconsistency and mismatch in the official data and the actual extent of the problem. Hate crime incidents are more common than official data suggest, and underreporting is a major problem (Dilmaç et al., 2021). There is no single authority collecting data on hate speech. The lack of data does not allow us to frame the extent of the problem and intervene effectively. Migrants, especially undocumented migrants, and refugees rarely report these incidents to law enforcement, due to distrust of law enforcement, lack of confidence in the impact of reporting, fear of victimization, and lack of awareness of their rights.

An effective data collection mechanism for recording hate speech is necessary to locate and evaluate the problem. Data collection should be in partnership with civil society organizations, not only from police records.

Although services are provided to support victims of hate speech, an effective prevention system would be beneficial for the communities, including awareness and education campaigns for youth and adults.

Finally, capacity-building and informative training on hate speech and discrimination should be provided to law enforcement agencies and CSOs in identifying, recording, preventing, and combatting hate speech and discriminatory practices.

BIBLIOGRAPHY

Bilewicz M. & Soral W., "Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization", Political Psychology, 2020, available at: https://psycnet.apa.org/record/2020-45123-001

Brown A., "How Social Media Monetization Is Evolving In The Face Of Algorithmic Bias:A Discussion With Nick McCandless", Forbes, 2021, available at: https://www.forbes.com/sites/anniebrown/2021/11/14/how-social-media-monetization-is-evolving-in-the-face-of-algorithmic-bias-a-discussion-with-nick-mccandless/

Bruce Lanner Smith, Propaganda, The Editors of Encyclopaedia Britannica, 21 January 2021

Castaño-Pulgarín et al., "Internet, social media and online hate speech. Systematic review. Aggression and Violent Behavior", 2021

Chaplinsky v. New Hampshire, 315 U.S. 568 (1942), available at: https://supreme.justia.com/cases/federal/us/315/568/

Cohen v. California, 403 U.S. 15 (1971), available at: https://supreme.justia.com/cases/federal/us/403/15/

Communist Party of the United States v. SACB, 367 U.S. 1 (1961), available at: https://supreme.justia.com/cases/federal/us/367/1/

Costa Pinto S. & Nascimento Cunha M., "NONVIOLENT COMMUNICATION – A LITERATURE REVIEW", August 2023, available at: https://www.researchgate.net/publication/373448238_NONVIOLENT_COMMUNICATION_-A_LITERATURE_REVIEW

Council of Europe (20/05/2022) - Recommendation CM/Rec(2022)16[1] of the Committee of Ministers to member States on combating hate speech, Council of Europe Portal, available at: https://www.coe.int/en/web/cyberviolence/-/council-of-europe-recommendation-cm-rec-2022-16-1-of-the- committee-of-ministers-to-member-states-on-combating-hate-speech#:~:text=This%20 Recommendation%20CM%2FRec(2022,within%20a%20human%20rights%20framework

Council of Europe, Bookmarks - A manual for combating hate speech through human rights education, 2020, available at: https://book.coe.int/en/human-rights-democratic-citizenship-and-interculturalism/8155-bookmarks-a-manual-for-combating-hate-speech-online-through-human-rights-education-2020-revised-edition.html

HOPE For Children

HATE TRACKERS BEYOND BORDERS

Agenzia Italiana per la Gioventù

Co-funded by the European Union

Council of the European Union (2008) - 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, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32008F0913

Council of the European Union, Assessment of the Code of Conduct on Hate Speech online State of Play, 2019, available at: https://commission.europa.eu/document/download/a5c92394-8e76-434a-9f3a-3a4977d399bb_en?filename=assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play_.pdf

Cowan G., Judgments of Hate Speech: The Effects of Target Group, Publicness, and Behavioral Responses of the Target, Journal of Applied Social Psychology, 1996

Cyprus's Constitution of 1960 with Amendments through 2013, 6th June 2017)

Dieudonné M'BALA M'BALA against France, Application no. 25239/13, ECHR, 2015)

Dilmaç J. A., Kocadal Ö., Tringides O., Public Discourse Of Hate Speech in Cyprus: Awareness, Policies and Prevention, Friedrich Ebert Stiftung, 2021, available at: https://library.fes.de/pdf-files/bueros/zypern/17404.pdf

Donzelli S., Countering Harmful Speech Online. (In)Effective strategies and the duty to counter speak, Bielefeld University, 2021 DovidioJ.

F., SchellhaasF. M. H.& Pearson R. A.,Prejudice, Oxford Research Encyclopedias,2019

European Commission (2016) – The EU Code of conduct on countering illegal hate speech online, available at: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

European Court of Human Rights (20 October 2015) – M'Bala M'Bala v. France, available at: https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-160358%22]}

Gelber K., Differentiating hate speech: a systemic discrimination approach, Critical Review of International Social and Political Philosophy, 24:4, 393-414, 2019, available at: https://philpapers.org/rec/GELDHS-2

Gelber, K., & McNamara, Evidencing the harms of hate speech. Social Identities, 22, 324-341, 2016, available at: https://psycnet.apa.org/record/2016-14217-008

Kathimerini, Archbishop's Christmas message with references to the Cyprus problem, 25/12/2017, available at: https://www.kathimerini.com.cy/gr/politiki/xristoygenniatiko-minyma-arxiepiskopoy-me-anafores-sto-kypriako

Keita G., "Discussing discrimination, American Psychological Association", 2016, available

at: https://www.apa.org/topics/racism-bias-discrimination/keita

KISA, "Cyprus: NGO Follow-up Report", 2017, available at: https://kisa.org.cy/wp-content/uploads/2017/05/KISA_submission_CERD_04-17_final.pdf

Law 134(I)/2011 of the Criminal Code "The Combating Certain Forms and Manifestations of Racism and Xenophobia through Criminal Law". (https://www.cylaw.org/nomoi/enop/non-ind/2011_1_134/full.html)

Law 84(I)/2003, Section 47(1)(b) of the Criminal Code "Actions against the sovereignty of the Republic, etc." (https://www.cylaw.org/nomoi/enop/ind/0_154/section-sc62405e78-42a0-481c-a3bd-372eeda41a7b.html)

Law 87(I)/2015 of the Penal Code, Section 99(A) "Incitement to violence or hatred base on sexual orientation or gender identity". (https://www.cylaw.org/nomoi/enop/ind/0_154/section-sc6b5f2a28-2aba-4385-6ce7-07d5b97c9d6f.html)

Matamoros-Fernández A., "Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube, Information, Communication & Society 20", 2017

Matamoros-Fernández A. & Farkas J., "Racism, Hate Speech and Social Media: A Systematic Review and Critique", 2020.

Mcleod S., "Stereotypes In Psychology: Theory & Examples, Simply Psychology", 2023, https://www.simplypsychology.org/katz-braly.html

Meta Transparency Center – Hate Speech, https://transparency.meta.com/it-it/policies/community-standards/hate-speech/

Mendel, T., "Freedom of Expression: A Guide to the Interpretation and Meaning of Article 10 of the European Convention on Human Rights, Centre for Law and Democracy", available at: https://rm.coe.int/16806f5bb3

Munt, R. S., "East European Feminisms" Feminist Encounters, 4(2), 2020, https://doi.org/10.20897/femenc/8509.

National Socialist Party of America v. Village of Skokie, 432 U.S. 43, 1977, available at: https://supreme.justia.com/cases/federal/us/432/43/

Newberry C., 2023 Facebook Algorithm: How to Get Your Content Seen, 2023

Official Gazette of the Republic of Cyprus No. 2682 of the 21 of February 1992, available at: https://www.cylaw.org/nomoi/arith/1992_3_011.pdf

Official Gazette of the Republic of Cyprus No. 566 of the 30 of March 1967

Oremus W. et al., "How Facebook shapes your feed", The Washington Post, 2021, available at: https://www.seattletimes.com/business/how-facebook-shapes-your-feed/

Paz M., Montero-Diaz J., Moreno-Delgado A., "Hate Speech: A Systematized Review", 2020, available at: https://www.researchgate.net/publication/346845790_Hate_Speech_A_Systematized_Review

Reynder D., Countering illegal hate speech online - 7th evaluation of the Code of Conduct, European Commission, 2022

TikTok Community Guidelines – Hate Speech and Hateful Behaviors, available at: https://www.tiktok.com/safety/en/countering-hate/

Travers M. "The Psychological Harms of 'Slacktivism,' According To A Psychologist", Forbes, October 31, 2023

UN General Assembly resolution 2106 (XX) (21 of December 1965), "International Convention on the Elimination of All Forms of Racial Discrimination"

United Nations Strategy and Plan of Action on Hate Speech, 2019, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml

University of Sussex, "Is digital activism effective?", July 2023, available at: https://study-online.sussex.ac.uk/news-and-events/social-media-and-campaigning-is-digital-activism-effective/

World Economic Forum, "Global Gap Gender Report", 2023, available at: https://www.weforum.org/publications/global-gender-gap-report-2023/

X Help Center, "Hateful Conduct", 2023, available at: https://help.x.com/en/rules-and-policies/hateful-conduct-policy#:~:text=April%202023,%2C%20disability%2C%20or%20serious%20disease.