universität wien

MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„Freedom of expression and hate speech in the case law of the European Court of Human Rights and in the practice of social media"

verfasst von / submitted by

Darya Novatorova, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Philosophie (Mag.phil.)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 841
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Magisterstudium Publizistik-u. Kommunikationswissenschaft
Betreut von / Supervisor:	Univ.-Prof. Dr. Katharine Sarikakis

Table of contents

1. Introduction

The collision between freedom of expression and hate speech is one of the most essential problems of the modern media law. On the one hand, freedom of expression is a fundamental human right and one of the main conditions for the existence of a truly democratic society. In the western countries intellectuals have emphasized this fact starting with the Modern history. After the Second World War the right for freedom of expression was repeatedly reiterated and secured in the major Conventions and Declarations of the 20th century, including the European Convention on Human Rights (1950). In the 21st century freedom of expression remains a crucial and an absolute component of human freedom and the bravest of us continue to fight for it all over the world.

Hate speech, on the other hand, is a destructive and offensive content – words, images or videos – that aims to encourage hatred and discrimination in the society. While rapid development of the Internet and new media, without doubt, has considerably contributed to the realization of freedom of expression worldwide, it has also facilitated the dissemination of hate speech, making it a global issue of our time. The key question is how to establish control over the latest, without violating the former: who and how determines, what hate speech is and how to combat it? Do different concerned parties – in particular, traditional courts and social media – have equal approaches towards hate speech and if not, how do they differ? What are the ways for combating hate speech that

social media have already developed, are they effective and what other alternatives could be employed? These are the underlying research questions of the current thesis.

As the European Court of Human Rights has acknowledged, no universally accepted definition of hate speech exists in modern media law, which, of course, further complicates combating this harmful phenomena (ECtHR, Factsheet hate speech, 2019). To our knowledge, one of the most precise definition of hate speech was given by Council of Europe who in 1997 defined it as "all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin" (Recommendation No. R 97(20), Appendix).

As will be demonstrated later, by no means this definition is the only one or universally accepted. Though it must be emphasized that already in this definition of Council of Europe it is possible to establish two necessary components that together constitute hate speech: (1) promotion/justification/dissemination of hatred based on (2) one of the many possible causes, whether it is nationality, ethnicity, migrant origin or affiliation with a minority group of any kind. Two further "causes" that for some reasons were not included in the definition of Council of Europe in 1997, are gender and religion. In fact, gendered hate speech and hate speech based on religious beliefs are a widespread issue of Internet era and will be discussed later in the first part of the study.

In an attempt to answer the research questions specified in the respective paragraph and taking into account the significance of the European Court of Human Rights for the European Convention on Human Rights[1] and, in general, for the European law practice, it was decided to focus in the present thesis on the recent case law of the European Court of Human Rights related to hate speech and to compare the Court's approach with such of the major social media – namely, Facebook, Twitter, and YouTube. The choice of these social media was dictated by their evident popularity, wide spreading in the society and influence over people's lives.

After the theoretical overview of the related issues in the first part of the study, in order to determine how the European Court of Human Rights separates protected speech that can "offend, shock or disturb" and hate speech that needs to be forbidden, a content analysis of the selected Court's cases was conducted in the second part of the thesis. The Court's definition of hate speech and proposed typology are also provided in this part of the study.

Accordingly, the third part of the thesis is dedicated to the practice of the above mentioned social media. Their current policies regarding hate speech, as well as the respective guidance for the users are analysed. Attempts of combating hate speech in the alternative ways are discussed in the paragraphs 5.3-5.4.

[1] The European Court of Human Rights was established in 1959 for protecting the fundamental human rights enshrined in the Convention. See: http://www.echr.coe.int/Pages/home.aspx?p=basictexts&c=#n1359128122487_pointer

Finally, key observations regarding the comparison between European Court of Human Rights' approach towards hate speech and that of the social media are provided in the Conclusion.

2. Theoretical background

2.1. Freedom of expression

There is no need to prove what essential role freedom of expression plays in any democratic society: it is its keystone and an indispensable condition. The right for freedom of expression is regarded as one of the most significant human rights in many countries, especially – in the states-members of the Council of Europe[2] whose main normative act – the European Convention on Human Rights – proclaims and guarantees freedom of expression in the well-known Article 10.

According to this Article, freedom of expression consists of two principal parts: (1) freedom to hold opinions and (2) to receive or to share information. Significantly, in the second part of the same Article 10 it is emphasised that freedom of expression is not absolute and may go hand in hand with responsibilities. The latest – possible restrictions – must be "prescribed by law" in the interest of a democratic society, protecting various concepts, but among others – public safety, order, health and morals, and even "the reputation or rights of others" (European Convention on Human Rights, Art. 10.2). Thus, it is evident, that the European Convention on Human Rights codifies freedom of expression while at the same time acknowledging its possible limits and leaving sufficient legal space for the prohibition of hate speech.

[2] Current countries-members of the Council of Europe can be found here http://www.coe.int/en/web/portal/47-members-states

The debates around freedom of expression versus hate speech mainly circles around one issue: how to maintain diversity of thoughts and views, while at the same time preserving dignity of all groups of citizens (Maitra, McGowan, 2018, p.305). The second part of the Article 10 of the European Convention on Human Rights – important and indispensable as it is – is very broad and therefore has caused many questions and interpretations by courts, as well as became an object of analysis for generations of media law researches (among many others, Korn, 2014; Holoubek, Kassai, Traimer, 2014). Provided by the Convention list of conditions, when the restrictions of freedom of expression are necessary in a democratic society, makes it impossible for the courts to use the Article 10.2 automatically. On the contrary, in each case courts have to consider all the circumstances and the details.

To make things even more complicated, when discussing freedom of expression in addition to the Convention it is necessary to take into account the case-law of the European Court of Human Rights. In the famous and historical case Handyside v. United Kingdom, 1976, the Court has stated that not only that information should be free that is considered inoffensive or indifferent, but also any ideas that can "offend, shock or disturb" any particular group of people (Handyside v. United Kingdom, 1976, par. 49). The Court motived it by "pluralism, tolerance and broadmindedness" that are the pillars for any democratic society (ibid).

For decades, this statement about the information and ideas that can "offend, shock or disturb" was a keystone of all modern media law practice. It establishes the broad

understanding of freedom of expression and also implies tolerance for some speech as a price for living in "a democratic society". As was summarized by McGonagle, democracy, too, has "rough edges" and "tough talk" should be regarded as a part of public debate (2013, p.5).

At the same time, the decision in Handyside v. United Kingdom does not abolish or contradict the second part of the Article 10 quoted earlier: there is still information that must be restricted when prescribed by law and "in the interests of a democratic society". In particular, it concerns hate speech.

2.2. Defining hate speech

As was mentioned in the Introduction, there is no universally accepted definition of the term "hate speech" (ECtHR, Factsheet hate speech, 2019). In fact, hate speech definitions vary by country, jurisdictions and particular type of law (D'Souza, Griffin, et al, 2019, p.943). One of the definitions that were found during the current research was provided by the European Union Agency for fundamental rights (FRA) which defines hate speech as "the incitement and encouragement of hatred" as well as "discrimination or hostility towards an individual" based on prejudice connected with any particular characteristic, whether it is "sexual orientation or gender identity" (Hate Speech and Hate Crimes Against LGBT Persons).

In comparison to the definition of hate speech given by the Council of Europe and quoted in the Introduction, the definition of the European Union Agency for fundamental rights can be regarded as incomplete, as it clearly excludes racial hate speech, or hate speech based on nationality, ethnicity, migrant origin etc. On the other hand, it is worse noticing the presence of the same two components that were highlighted in the definition of the Council of Europe: (1) encouragement of hatred (2) based on one or another characteristic of a victim.

Unfortunately, in practice defining hate speech is not that straightforward. As was observed by Parekh, in different countries different speech is considered to be hate speech (Parekh, 2012, p.40). In some instances this speech expresses certain views, but does not call for actions; in another the speech under consideration is abusive but not threatening; in some situations the speaker expresses dislike of a certain group of people, but not hatred, and wishes no harm (ibid). Parekh emphasizes that this confusion is a direct result of the tendency to summarize under the concept of "hate speech" all kinds of uncivil or hurtful speech (ibid).

Parekh claims that it is more reasonable to define hate speech as incitement of hatred against a group of people, based on their nationality, sexual orientation, religion, gender, etc. In other words, the researcher supports the definition proposed by the Council of Europe and quoted in the Introduction. However, according to him, the concept of hate speech consists primary of three, not two components:

(1) It is directed against concrete individual or group of people based on a particular feature;

(2) It stigmatizes this individual or group of people by prescribing them highly undesirable qualities and

(3) It utilises these negative qualities to make this individual or group "a legitimate object of hostility" (Parekh, 2012, p. 41). Hate speech does not necessarily results in violence and must not necessarily be expressed with offensive of abusive language (ibid).

Gelber and McNamara also admit that, whereas the term "hate speech" is widely used, it does not have a single meaning (2016, p.324). For their own study they adopted the three-component definition of Parekh.

Richardson-Self, while reiterating the fact that there is no one accepted definition of hate speech among scholars, underlines that it is speech that a) is hostile and b)silences, humiliates, threatens or discriminates and c) targets group of people based on common traits, such as race, religion beliefs, disability, sexual orientation or gender (2017, p.256). Hate speech per se in an oppressive act, argues Richardson-Self (ibid, p.257).

Yong, too, defines hate speech as speech that attacks people based on their particular characters, whether it is race, nationality, gender, sexual orientation, religion believes or any other identity marker (2011, p.386). Yet, the researcher shares the view that the term "hate speech" is "unsatisfactory" (ibid). Yong debates that different kinds of speech are

understood under this term and that appropriate response to them should differ from case to case (ibid).

Moreover, providing a definition to hate speech in practice becomes even more difficult in the countries which experience great social tensions or had especially troubled historical past. For example, hate speech is a relatively new term in South Africa, where racism and discrimination until recently were institutionalized in the apartheid policy (Davids, 2018, p.297). It is worth noticing that now hate speech is criminalized in the country, but nevertheless it experiences a significant increase in both hate speech and hate crimes (ibid).

Finally, there is a view among some scholars that to create one final definition to hate speech is difficult or even not possible because hate speech is a continuum (D'Souza, Griffin, et al, 2019, p.958). Whereas in this case the researchers initially concluded it in relation to gendered hate speech, the idea of hate speech as continuum or spectrum (McGonagle, 2013, p.4) seems to reflect the nature of this phenomenon. In particular, it could explain not only the problem of definition, but also different approaches and debates regarding hate speech regulation.

Nevertheless, despite the lack of one single definition, it is possible to observe certain similarities in the proposed ones. To sum it up, we propose to understand under hate speech, utterances that (1) encourage hatred (2) towards individuals or groups of people (3) based on various particular group characteristics (4) which are stigmatized and (5) are

employed to legitimize hostility. Furthermore, hate speech (6) may be understood as a continuum that (7) does not necessarily result in violence, but (8) silences, discriminates and threatens targeted individuals or groups.

2.3. Hate speech and hate crimes

Scholars often acknowledge that hate speech involves incitement of hatred (D'Souza, Griffin, et al, 2019, p.945), and thus the concept of "hate speech" is sometimes misunderstood as a synonym to the notion of "hate crime". Whereas these concepts are related, it is significant for the purpose of the current thesis to underline the difference between them.

Hate crimes can be regarded as violent manifestations of intolerance (OSCE, 2009, p.11) that usually have an impact not only on the victim, but also on the whole group to which the victim belongs. From other crimes hate crimes can be distinguished by bias motive. It is important to understand that hate crime can be manifested as an intimidation, threats, assaults, murder or similar criminal offence (ibid, p.16).

The bias motive and criminal offence which have occurred are two principal features of the hate crimes. It means that a perpetrator selects a victim not randomly, but on the basis of his or her belonging to a special group. Thus, the key issue is *what* the victim

represents, and not *who* he or she is. (ibid, p.17). Race, language, religion, gender and similar aspects are possible examples of basis of the hate crimes (ibid, p.9).

On the other hand, hate speech will still be penalized in some countries even if no criminal base offence have occurred, but only speech that promotes hatred or insults certain group of people (ibid, p.25). For example, if someone violates a person's or nation's honor or dignity, it will be hate speech, but not a hate crime.

In other words, the term "hate crime" is used for acts, whereas the concept of "hate speech" is related to discriminatory views/speech (ibid, p.17). Thus, specific content – words, images, videos – can be regarded as a principal future of hate speech, but not that of hate crime. At the same time, hate speech can easily lead to hate crime. For instance, Davids argues that this is what happens in Charlottesville in August 2017 when a rally of supremacists and nationalists evolved into violence (2018, p. 298). On the other side of the globe, in Australia gendered hate speech fuels gender-based violence, which is a view of government bodies and some scholars (D'Souza, Griffin, et al., 2019, p.940). Yet, there is a notable absence of laws regulating gendered hate speech in the country (ibid).

OSCE believes that legislation is only a part of the possible answer to the problem of hate crimes (OSCE, 2009, p.11). In order to combat hate crimes, there is a need in a comprehensive national programs which should also include education, special training, accurate data collecting, etc. (ibid, p.12). On the other hand, there are extreme variations

between hate speech laws in different countries, which can be explained by different constitutional and even philosophical approaches (ibid, p.26).

2.4. Hate speech: contextual typology

Whereas one universally accepted typology of hate speech does not exist, it is possible to establish such contextually based on the most common themes that are to be found in hate speech examples, namely:

1) Hate speech based on nationality, race or ethnicity, including hate speech against people with migrant origin and incitement to genocide

Racist hate speech exists in all parts of the world and constitutes a critical challenge for human rights (UN, General Recommendation 35, 2013, p.10). Racist hate speech does not necessarily equals to explicitly racial comments but may use indirect language (ibid, p.3). Among others, common targeted groups of this type of hate speech are indigenous peoples and people with migrant background including refugees and migrant workers (ibid, p.2). Davids debates that racist hate speech not only results in violence, but also leads to discrimination of a particular group of people (2018, p.298).

According to Maitra and McGowan, regulation of racist hate speech should not involve any concerns related to freedom of speech (2010, p.370). In turn, while discussing

approaches towards racist hate speech and its qualification as an offence punishable by law, UN Committee on the Elimination of Racial Discrimination proposed to take into account several contextual factors, such as

- content and form of a particular commentary (for instance, whether it is provocative or not);
- existent economic, political and social climate (whether there is already a discrimination against a targeted group or not);
- potential influence of the speaker and his/her audience;
- potential reach of hate speech;
- the objectives of it (UN, General Recommendation 35, 2013, p.5).

Although Committee on the Elimination of Racial Discrimination introduced this contextualization to the racist hate speech only, analysis of contextual factors might be useful for approaching other types of hate speech as well.

2) Political hate speech, including anti-Semitist speeches or Holocaust denial

Political expression and hate speech related to political issues should enjoy different legal protection, but in practice it is often difficult to distinguish between them, claims McGonagle (2013, p.19). In case of ambiguity, the researcher proposes to analyse not only the content of the speech itself, but also its form, context, position of a speaker in a

society and aim of speech (ibid). For instance, a clear separation should be made between speech that aims at contributing to public debate and speech that incites hatred (ibid).

Understandably, freedom of political expression should be protected in a democratic society. However, this freedom is not absolute, emphasises McGonagle, but involves responsibilities for the speaker (ibid, p.20). As will be demonstrated in the second part of this thesis, protecting political expressions while simultaneously prohibiting political hate speech is a difficult challenge even for the European Court on Human Rights. This collision between political expression and political hate speech may become even more acute online (ibid).

From the legal perspective, instances of Holocaust denial are a more clear issue. European Court of Human Rights determined in the case Garaudy v. France that the mere questioning of the Holocaust constitute an incitement to hatred against Jews and as such, contradicts the values of the Convention (2003, par. 23).

3) Hate speech based on religious beliefs or lack of such

One could assume that religious hate speech involves discrimination based on religious believes, but Bonotti emphasises that the relation between hate speech and religion is not that straightforward (2017, p.259). Firstly, there is hate speech directed against religious minorities. Secondly, there are religious people who attempt to justify their hate speech

against particular groups of people based on, for instance, their sexual orientation, with an argument of religious freedom (ibid). In fact, they utilize hate speech as a means of propaganda of their faith (ibid, p.272).

In this case, maintains Bonotti, religious believers spreading hate speech must be treated as any other citizens (ibid). The argument about religious freedom cannot be taken into account because there are other ways to spread their religious message, as well as express their disapproval of other people's lifestyles (ibid).

4) Gendered hate speech

Analysis of gendered hate speech is rare, argues Richardson-Self who believes that separate study of this kind of hate speech is required (2017, p.257). Interestingly enough, the researcher makes a distinction among (a) misogynistic speech that she considers to be hate speech, and (b) oppressive sexist speech that in her opinion does not amount to hate speech, but can be, nevertheless, violent (ibid). On the other hand, speaking about misogynistic speech versus sexist speech, D'Souza, Griffin, et al. defend their choice of the term "gendered hate speech", emphasizing that it is an ongoing construction of gender that involves using gendered norms for attacks on people or groups of people (2019, p.955). Yet, D'Souza, Griffin, et al. agree that language maintains stereotypes that exist in society about gender, and thus gendered hate speech can be regarded as a means of

spreading misogynist hostility in order to support patriarchal structures in a society (ibid, p.967).

Explaining the difference between misogynistic speech and oppressive sexist speech, Richardson-Self clarifies that misogyny involves hostility against women, whereas sexism is an instrument of reinforcement or justification of patriarchy (2017, p.261). Speech that belittles or patronizing women – for instance, calling grown women "girls" – can be regarded as morally unaccepted and sexist, though it is not hate speech (ibid, p.262). On the other hand, misogynistic hate speech usually contains direct hatred towards women which often manifests itself through rape-threats or comments (ibid, p.265).

In contrast to other types of hate speech, misogynistic utterances usually target not a single group – that is, not all women – but only women who do not comply with patriarchal standards. Richardson-Self calls this phenomena "intradivisional speech" (ibid, p.267). Yet, this kind of speech should still be regarded as hate speech as it degrades and dehumanizes targeted people as much as any other type of hate speech (ibid, p.268).

Other scholars maintain that gendered hate speech must be analyzed in socio-political context where it enforces patriarchal structures and norms (D'Souza, Griffin, et al, 2019, p.940). Gendered hate speech is inextricably connected with traditional gender boundaries which results in women being target of it, for instance, for voicing their ideas

or participating in society's life (ibid, p. 956). Nevertheless, whereas women and girls are much more likely to become victims of gendered hate speech, it should be noted that it is not necessarily directed at females only: men, too, can become victims of it if they do not comply with traditional roles (ibid).

Interestingly enough, some scholars further distinguish so-called "conservative" definition of gendered hate speech and "progressive" one (ibid, p. 958). They argue that the former focuses on protecting public order, whereas the latest – on protecting the victims (ibid). In other words, supporters of the "conservative" view on gendered hate speech prohibit it because it may cause criminal actions, whereas supporters of the "progressive" view simply ban gendered hate speech because of its harm to targeted individuals (ibid, p.959).

5) Hate speech based on sexual orientation

European Court of Human Rights first recognised homophobic hate speech in 2012, in the case Vejdeland & Others v. Sweden (McGonagle, 2013, p.12). The Court did not provide a full definition to it, but stated that homophobic hate speech is directed against a person of group of people based on their sexual oritnation (Vejdeland & Others v. Sweden, 2012, par.42). The Court debated that despite widespread of homophobic hate speech, there were no established standards for approaching the problem in 2012 (ibid).

As it is often the case with hate speech, the boundaries between hate speech based on sexual orientation and gendered hate speech can be blurred. Thus, lesbians are often attacked on both grounds simultaneously which supports the assumption that hate speech is employed to control gendered boundaries and behaviour (D'Souza, Griffin, et al, 2019, p.955).

Overall trend however is that homosexual or bisexual people are more likely to become targets of hate speech involving death or rape threats than their heterosexual contemporaries (ibid, p.957).

2.5. Other categorisations of hate speech

Of course, it is necessary to underline that contextual categorisation presented above is not absolute. As will be demonstrated in the case law analysis later, quite often these categories intermingle and hate speech instance may be attributed to more than one of them.

An alternative typology of hate speech can be based not on the content of speech itself, but on the way it is disseminated. The first category of hate speech then would be “face-to-face encounters”, the second – generally circulated hate speech (Gelber, McNamara, 2016, p.325). Whereas the distinction between these two types of hate speech is not always obvious, regulation of the second is much more controversial than the latest (ibid,

p. 325). Yet, based on the interviews with people, targeted by hate speech, scholars Gelber and McNamara concluded that these two mentioned types were not experienced different by victims, whether by "seriousness" or by "harmfulness" (ibid, p. 336).

Finally, yet another completely different categorisation of hate speech was given by Yong who distinguished four different categories (2011, p.386):

1) "Targeted vilification", or speech that has a main intention to intimidate, to wound or to insult targeted individuals or groups (ibid, p.394). Targeted hate speech may occur either through face-to-face encounters or without any direct contact, but its main characteristic is that it is "narrowly directed" at the victims (ibid). This category of hate speech, according to the researcher, should not be covered by the principle of free speech as it is hard to imagine that expressing hostility or contempt can lead to personal development of the speaker or correspond to any other interests that are supported by the principle of free speech (ibid, p.395);

2) "Diffuse vilification" that is covered by this principle but must not be protected (ibid, p.386, p.396). Yong understands under this category hate speech that attempts to intimidate, insult or wound, but is not directed at particular individuals or groups of people and is addressed to a wide audience of potentially sympathetic listeners – for example, speeches during Nazi marches (ibid). Yong argues that this kind of speech constitutes somewhat "greater degree of free speech interests" than targeted vilification, but contextually significance of such speech is usually low because the main aim of

diffuse vilification is to intimidate, not to disseminate any ideas (ibid, p.397). In other words, it is a low-value speech and its regulation is necessary for prevention of intimidation (ibid, p.398);

3) "Political advocacy", or speech that propagates exclusionary or eliminationist policies and that, too, argues Yong, is covered by the principle of free speech, but must not be protected (ibid, p.386). It must be emphasized that under "exclusionary policy" Yong understands "exclusion from full citizenship" of particular groups of people based on race or religion believes, and under "eliminationist policies" – "ethnic cleansing" or "forced repatriation" of these groups (ibid, p.398). As such political advocacy can result in enactment of these principles, Yong claims that this kind of hate speech, too, must be unprotected by free speech doctrine (ibid);

4) Other instances of speech that presents "adverse judgements" about a particular group of people based on their race or religion. According to the researcher, this category of speech must be protected by the principle of free speech even despite its potential harm (ibid, p.386, p.401). Explaining his position, Yong claims that this kind of speech is assertions or opinions and thus it can be answered through counterarguments (ibid).

In other words, Yong suggests that speech that is either uncovered (1) or unprotected (2-3) is "regulable" and may be regulated (ibid, p.388).

2.6. Philosophical and practical reasons to ban hate speech

Before providing an overview of legal base for hate speech regulation, it is necessary to determine why hate speech must be regulated. Whereas this issue seems to be straightforward – hate speech unlawfully threatens targeted people or groups of people and violets their dignity – scholars argue that the consequences of hate speech are much more diverse and long-term, and eventually can affect the whole society.

Thus, Parekh emphasized that hate speech is harmful for public debate, as it negatively affects communities' moral sensibility and distorts mutual respect in society (2012, p.44). Hate speech creates the environment of hostility and fear and intimidates the target group so that these people find themselves unable to participate in public life (ibid). Of course, simultaneously it affects personal lives of targeted people as they feel themselves humiliated and discriminated against (ibid).

Similar position support Gelber and McNamara, who underline that hate speech can be regarded as "an existential attack" on a person's dignity and its negative effects are long-term and enduring (2016, p.325). Researchers argue that two types of harms of hate speech are usually distinguished:

1) constitutive, or harm resulted from saying hate speech per se, and

2) consequential, or harm that can be a result of hate speech (ibid). For instance, hate speech can lead to negative stereotypes and create an environment where there stereotypes are normalized (ibid).

While contacting interviews with the people in Australia, targeted by hate speech, Gelber and McNamara observed that they acknowledge being hurt, upset and frightened (ibid, p.328, p.333). In a number of cases interviewers reported that fear prevented them expressing themselves or to act against hate speech (ibid, p.333). Even silence and withdrawal from social life was mentioned by interviewers as a means to avoid hate speech (ibid, p.334), which eventually supports the view of Parekh regarding the exclusion of targeted groups from social life. Similar ideas were expressed by Davids who argues that racist hate speech results in exclusion (2018, p.298). The scholar emphasizes that peaceful coexistence of people in a society is not possible in a climate of hate speech (ibid, p.306). Finally, D'Souza, Griffin, et al. acknowledge that one specific type of hate speech – namely, gendered hate speech – indirectly contribute to perpetuating gender-based violence (2019, p.963).

Yet not everyone agrees with the idea that hate speech must be banned. American scholar Baker, who defines himself as a supporter of "almost absolute protection" of freedom of speech (2012, p. 57), provides two arguments why prohibition of hate speech is not an ultimate solution of the problem. Firstly, he argues that prohibition will not effectively reduce potential harm of such speech (ibid, p.72). Secondly, he believes that such measures will, on the contrary, aggravate the existent problems in society without properly addressing them (ibid).

Baker claims that direct response and open criticism of, for instance, racist views, can be more effective than prohibition of hate speech (ibid). Law regulation, on the other hand, can cause hate speech to "go underground" so that the real extent of the issue not be evident (ibid, p.73). Moreover, Baker believes that prohibition may enrage people of groups of people who support racist views, thus encouraging them to act (ibid). Conflicts within a society must be approached as political, not "violent struggle" (ibid, p.74). Politicians must address underling origins of racism and not hate speech that results from it (ibid).

While Baker's arguments evidently have some basis, one could retort that there is hate speech that cannot be answered in public debate, simply because it is motivated by pure hatred of troubled individuals and thus is not reasonable or logical arguments to be discussed in a dialogue. Furthermore, as was discussed above, there are various types of hate speech and it seems that the existence in a society of xenophobia, religious disagreements, misogyny and other forms of tensions does not need proof in the form of hate speech. However, open criticism, mentioned by Baker, as well as general condemnation of hate speech by the society, seem to be essential in combating this hurtful phenomena, especially in the internet era, where hate speech resistance cannot completely depend upon law enforcement.

Being an American scholar, Baker represents "permissive" American approach which will be discussed below and which differs from the European perspective that is the primary focus of the current thesis.

2.7. International legal base for freedom of expression and hate speech' regulation

Dealing with freedom of expression and hate speech and its regulation, it is necessary to establish the general European legal base of the question. Freedom of expression as a standard was first prescribed in 1948, in the Article 19 of the Universal Declaration of Human Rights. Two years later the European Convention on Human Rights, main normative act of the Council of Europe, guaranteed freedom of expression in the quoted above Article 10 and listed possible restrictions to it in the second part the same Article.

In 1966, Article 19 of the International Covenant on Civil and Political Rights reiterated freedom of expression and everyone's right to hold opinions without interference. At the same time, Article 20 of the Covenant proclaimed that "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law" (International Covenant on Civil and Political Rights, 1966, Article 19-20).

Similar approach can be found in the Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination (1965) which disapprove propaganda and organizations that advocate for superiority based on ethnicity or skin colour, or justify discrimination or racial hatred (International Convention on the Elimination of All Forms of Racial Discrimination, Article 4). However, practical interpretation of these

Conventions differs from state to state (OSCE, Preventing and responding to hate crimes, 2009, p.54).

Additionally, incitement to genocide that is also amount to hate speech (McGonagle, 2013, p.6) is forbidden by the Article 3 of the Convention on the Prevention and Punishment of the Crime of Genocide which entered into force in 1951.

2.8. Combating hate speech: minimal and extensive regulation approaches

The widespread of hate speech has become a global issue due to the Internet and modern technologies. As mentioned in the Introduction, this issue, can be divided into two parts: hate speech's impact on the society and individuals, its contribution to the hate crimes and violation of human rights that is causes, and, on the other hand, the possible consequences of the hate speech's regulations to the freedom of expression. According to the former OSCE Representative on Freedom of the Media Miklós Haraszti, the general trend is to extend regulation of hate speech, rather than balancing it by "more reasoned speech" (Haraszti, 2012, p.xiii). The same view supports Bhikhu Parekh, who claims that exact forms of expression that are prohibited vary from country to country, but universal trend is observable even for the past few decades (Parekh, 2012, p.37).

With regard to hate speech's regulation, there are two possible approaches: the first one is a so-called "minimal regulation approach" and the second – "an extensive regulation".

The idea of the minimal regulation approach is based on the assumption that actual instigations to actual hate crimes must be criminalized, whereas simply offensive speech should be handled by dialogue in the press, courts, ethics organizations (Haraszti, 2012, xiii). Defenders of free expression often argue that, regardless of the disseminated content, freedom of speech will ultimately benefit the whole society (Downs, Cowan, 2012, p.1354). Respectively, proponents of the extensive regulation approach propose new speech bans into national criminal codes (Haraszti, 2012, xiii).

Already in 2012 Haraszti underlined that extensive regulation approach may put at risk "the very existence of an international human rights standards for handling hate speech" (ibid). The real consequences of such an approach in practice can be visible when analyzing the situation with freedom of speech in the post-soviet countries, where war on terror was twisted into "a fight against "extremism", a vague term that quickly "expanded to encompass almost all forms of political dissent" (Richter, 2012, p.290).

It is necessary to underline the word "vague" in this quotation: for instance, vague and too broad definition of the term "extremism" in Russian and other post-soviet countries' law has given the officials a possibility to "stifle debate in the media of issues of increasing public interest – such as the motivations underlying insurgency" (ibid, p.305). In other words, post-soviet authorities use anti-extremist laws to suppress and to control any possible political opponents, but they are not the only ones. Similar strategy is employed in other parts of the world, for example, in South Africa, where the same vagueness of the national laws against hate speech concerns scholars who believe it can

be used against any unpopular, critical or politically unwanted speech (Davids, 2018, p.298).

To summarize this view, it is a lack of clear and concrete definition of extremism, which was and is misused by the authoritarian governments for the restricting of freedom of expression. Taking into account the lack of a concrete definition of hate speech in international law, it is possible to assume that hate speech regulation may also be misused by other governments or political parties in their interests. This danger is even more evident if we consider that, as will be demonstrated below, social media to a certain extent replaced traditional courts while taking their own decision as to what information to permit and what to prohibit.

Of course, opponents of this view believe that freedom of expression is used in defense of hate speech. For instance, Hate Speech International[3] – an initiative that is dedicated to combating extremism and supporting journalists that cover this topic – estimates that whereas hate speech leads to political violence, "a liberal paradox" is created by supporters of freedom of expression who insist on its protection (Hate Speech International, Political backdrop). This view is also supported by Parekh who argues that freedom of speech protects arguments, ideas and opinions, whereas hate speech weakens democracy by spreading irrational fears and creating a sense of insecurity among its

[3] Hate Speech International monitors hate speech, as well as hate crimes and extremism worldwide, thus being a very useful tool for researchers in this area.

victims (2012, p. 48). Maitra and McGowan, too, underline that harm caused by hate speech excludes it from any protection under free speech principle (2010, p.364).

Prominent communication scholar Peter Molnar attempted to find a compromise between the two views. According to Molnar, the debate between proponents of minimal and extensive regulation approach is based on lack of information and generalization that result in misunderstanding (2012, p. 183). Molnar argues that supporters of freedom of expression do not sacrifice everything for the sake of this freedom, but rather have different opinions as to what is the most effective response to hate speech (ibid, p.184).

Molnar himself considers art and education "in the broadest sense", as well as an application of the "imminent danger test" to be the best response to hate speech in truly democratic countries (ibid, p.185). This test is based not on the content, but rather on the context of hate speech and even on the social environment: if a certain expression constitutes "a clear and present danger of violent action", it must be prohibited (ibid, p 197). If, however, no imminent threat of violence can be assumed, prejudice and hatred are better to counteract with cultural activities, not law enforcement, claims Molnar (ibid).

2.9. European and American approaches towards hate speech

Some scholars believe that support of minimal and extensive regulation may be explained by historical background, providing as example different legal approaches in European countries and in the USA. In particular, the difference between European and American approaches is evident when it concerns racist hate speech (Bleich, 2014, p.284).

Discussing this difference, Greene argues that European approach towards hate speech regulation is based on the idea that laws must protect minorities from democracy that went wrong (2012, p.92). On the contrary, American approach assumes that democracy required that all voices must be heard and self-expression should not be limited (ibid). As a result, there is "restrictive" European policy towards hate speech, and "permissive" American (ibid, p.94). According to the Supreme Court of the USA, American Constitution prevents adaptation of laws that are required by both Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination and Article 20 of the International Covenant on Civil and Political Rights, discussed earlier (ibid, p.95). In both cases the Conventions were ratified in the USA with some reservations.

Abrams, too, notices that there is a constitutional protection of hate speech in the USA, which is ultimately based on the famous First Amendment (2012, p.116). Abrams argues that hate speech prohibition would be a restriction based on the content, which is strictly against the principles of the First Amendment (ibid, p.118). The same idea is supported

by Bleich, who argues that in the USA it is almost impossible to convict someone of racist hate speech, unless this speech leads to “immediate violence” or is a direct threat (2016, p.284).

Explaining the difference between European and American approach, Abrams suggests that different historical background – European experience with Holocaust and Nazi versus American history – can lead to different decisions regarding hate speech regulation (2012, p.119). The same view shares Schauer who underlines that Holocaust denial is regarded unlawful in Austria, Germany, France and some other European and non-European countries, but not in the USA where such prohibition would contradict freedom of speech (2012, p.130). In France, however, laws against Holocaust denial, as well as laws against racist hate speech, are used as a legal base against expressed hostility (Suk, 2012, p.148).

Some historical background of discrepancy between European and American approaches is given by Bleich who believes the two positions started to evolve differently in 1960th (2016, p.284). According to the researcher, after the European Court on Human Rights was established in 1959 to enforce the Convention, it obtained an ultimate authority in Europe for drawing the line between freedom of expression and hate speech (ibid). In the USA such power since 1920th belonged to the Supreme Court. However, Bleich stops here, noticing that there is no theoretically plausible explanation of the difference between European and American approaches – only various hypotheses (ibid, p.285).

For example, Bleich mentions different "institutional strength of the courts", variations in political cultures, divergent legal base for protection of freedom of expression, generally diverse jurisprudence systems and even contrasting interpretations of legal texts made by individual judges (ibid). At the same time Bleich emphasized that also European and American approaches towards freedom of expression and hate speech are "analytically distinct", nevertheless they do not exclude each other (ibid, p.286).

Interesting idea related to the American approach towards hate speech regulation is expressed by Downs and Cowan who argue that free-speech defenders in the USA acknowledge harm of certain speech, such as Holocaust denial or Nazi's views, but yet consider freedom of expression to be more important (2012, p.1354). Downs and Cowan observe that it is easy to defend someone with whom you agree, but not so – to defend your opponent's right to speak (ibid). This being said, researchers remind that hate speech can be harmful and humiliating to individuals, once again underlining the complex issue of finding a balance between freedom of speech and hate speech regulation (ibid).

2.10 . Online hate speech: general overview

While providing unprecedented opportunities for communication and spread of knowledge, Internet facilitated "multi-directional" communication between its users (McGonagle, 2013, p.28). Lack of barriers that existed in traditional institutionalized

media increased opportunities to participate in public discourse, but also simplified dissemination of hate speech (ibid).

This observation is supported by the Council of Europe that called Internet "a turbo accelerator" of hate speech" (Council of Europe, 2013, p.1). According to Council of Europe, there are two principal reasons for this phenomenon: wide access, as well as anonymity which excludes responsibility that should accompany freedom of expression (ibid).

One could argue that anonymity also has its advantages when it comes to political debates, but indeed, the Internet gave rise to additional instruments for dissemination of hate speech. Nowadays online hate speech can be found not only on the websites, but is more common in the social media where this problem has different perspective and could potentially lead to grave consequences due to real names and personal information, which people often provide about themselves.

A special aspect of online hate speech is difficulty of combating it due to the speed of dissemination, as well as lack of training on the technical issues which sometimes prevent quick investigation and identification who is responsible for dissemination of hate speech (OSCE, Preventing and responding to hate crimes, 2009, p.56). McGonagle debates that other complication for online hate speech regulation is jurisdictional issue as different countries have different laws regarding hate speech which, in turn, influence policies of internet-providers (2013, p.30). The researcher claims that websites that disseminate hate

speech are purposely being hosted in countries with "favorable jurisdiction" for hate speech (ibid, p.29).

During the last decade various social media encountered hate speech on their portals and determined to (or were made to) address this issue in their guidelines and rules for users. Their policies, along with those of the internet-providers, are not unified as well as their automatic algorithms of content adjustment are not fully transparent. Being private actors, their involvement in online speech regulation could potentially lead to censorship, rightly warns McGonagle (2013, p.30). Yet the regulation of online hate speech is de-facto left to these private social media.

Their related policies will be analyzed in detail in the third part of the thesis.

2.11. Alternative ways to combat hate speech

Not only Molnar, but other scholars and organizations believe that apart from minimal or extensive regulation discussed above, there are other, secondary activities that can contribute to combating hate speech. Possible list of these activities is given in the resource guide for NGOs "Preventing and responding to hate crimes" prepared by the OSCE's Office for Democratic Institutions and Human Rights in 2009. OSCE acknowledges that there is a strong connection between hate crimes and hate speech and propose following ways to counteract the latest:

- raise awareness about the problem;

- monitor the hate content and advocate for the removal of hate speech from the Internet;

- conduct education activities, more precisely – education of the problem of cyberhate for teachers, students, parents, law enforcement;

- denounce and challenge the arguments or claims, hold politicians and public figures accountable before public opinion;

- and, finally, take legal actions if hate speech «crosses the threshold into crime»

(OSCE, Preventing and responding to hate crimes, 2009, p.53-56).

Similar idea about a need of alternative responses to hate speech was expressed in General Recommendation No. 35 of the UN Committee on the Elimination of Racial Discrimination, titled "Combating racist hate speech" (2013). In particular, this Recommendation underlines a necessity of educational, information and cultural strategies for combating hate speech (2013, p.9).

In turn, Council of Europe, too, implements a number of strategies to combat hate speech. Among others, Council of Europe proposes to deny legal protection to hate speech, provide access to the means of expression to minorities and advocate for an intercultural dialogue (McGonagle, 2013, p.8).

In must be noticed that measures proposed by OSCE, UN and Council of Europe should be regarded as a complex activity. One could assume, however, that these and other

possible measures should be viewed as additional and cannot fully substitute law intervention because it reduces hate speech, sends a powerful message and prevents hostility against targeted individuals or groups of people (Parekh, 2012, p.46).

Finally, in the academic literature on hate speech exists an opinion, that yet another way – or additional way – to combat hate speech is to provide a voice to the groups targeted with hate speech (Gelber, McNamara, 2016, p.326). Whereas this hardly can be regarded as the only needed measure, it could definitely influence the debate around hate speech, inform society's perception of it and, eventually, provide some support to the victims.

3. Research questions and methodology

As was briefly outlined in the Introduction, the main research question of this thesis can be formulated as "What is understood under hate speech in a court's practice and in the practice of social media: do these parties have equal approaches towards hate speech and if not, what differences can be observed?"

Additionally, several sub questions were formulated that are supposed to deepen an understanding of hate speech phenomena and illuminate possible discrepancy between courts' approach and that of the social media:

1) What arguments does a court use to determine what is hate speech and what is offensive speech that nevertheless must be protected for the sake of freedom of expression?

2) What types of hate speech can be found in a court's practice?

3) How do major social media define hate speech?

4) How do major social media attempt to combat hate speech?

5) Are these methods effective and sufficient or not?

6) What alternatives ways to combat hate speech are already exist or are being developed?

7) How does social media's practice regarding hate speech correspond to that of a court?

8) Do social media de facto replace a judge when controlling online hate speech?

9) What implications it could potentially have for freedom of expression?

As the scope of the thesis did not permit us to analyse case law of a number of courts, it was chosen to focus on the practice of only one court, namely, the European Court of Human Rights, established in 1959. The choice of this particular court is based on its significance for the European law practice and for the European Convention on Human Rights which norms it is supposed to protect, including freedom of expression, guaranteed by Article 10 of the Convention. The practice of the European Court of Human Rights related to hate speech is analyzed and compared with guidances and actions of the major social media, such as Facebook, Twitter, and YouTube that were selected due to their popularity and widespread in society's life.

Thus, according to the portal of web-statistic Statista.com, in April 2019 Facebook had an active monthly audience of 2 320 millions of users, YouTube – 1 900 millions, and Twitter – 330 millions (see: Statista, Most popular social media worldwide). Policies of a number of other popular social media were excluded from the research for various reasons. For instance, WhatsApp was excluded as a messenger, Instagram – as a portal, primary orientated on visual communication, Chinese social media platforms such as me WeChat, QQ, QZone and Sina Weibo were not analysed due to the linguistic inaccessibility.

The main method used in the current thesis was qualitative content analysis that is based on text interpretation (Mayring, 2000, p.1). It must be noted that content analysis as a research method has a long history in social sciences. It was defined as a systematic technique that involves compressing texts into content categories already by Berelson in

1952. Holsti in 1969 described content analysis as a research technique that "objectively and systematically" determines particular characteristics of a text (1969, p.14). Krippendorff already in 1980 established that content analysis involves collecting a set of texts, systematic reading of them, noticing consistent features and making assumptions about their meaning (p.3). These systematic inferences are supposed to answer the research questions of the study (ibid, p.25).

Stemler, too, emphasized that content analysis relies on categorization of data and coding (2001, p.3), with Weber earlier explaining that under a category is understood a group with "similar meaning or connotations" (Weber, 1990, p.37).

At the same time, qualitative approach aims at understanding of a particular phenomena (Vaismoradi, Turunen, 2013, p. 398). Qualitative content analysis suggests that material should be analyzed gradually and by dividing it into analytical units (Mayring, 2000, p.3). The same idea was expressed by Sparker who suggested dividing content into small units for a description (2005).

Content analysis is applied by legal scholars, too, who conduct a content analysis of legal texts and judicial opinions for scientific understanding of the law as such (Hall &Wright, 2008, p.64). In fact, one of the first content analysis of the legal practice was conducted by political scientist Fred Kort in 1957 who recorded and categorized facts from a number of judgments in an attempt to predict results of similar cases (ibid, p.67). Other scientists soon began to use content analysis of the judicial opinions in order to

understand judicial behavior (ibid, p.68). In the 1990 content analysis of legal practice became widespread among legal scholars in the USA (ibid, p.70).

Hall and Wright underline that content analysis can provide different information to the legal researches. For example, it may be useful for counting court results (ibid, p.85) or evaluating legal doctrine (ibid, p.87). They further theorize that exploration of legal doctrine consists of understanding of judicial reasoning (ibid, p.87), conducting normative legal analysis (ibid, p.88) or studying the landscape of case law (ibid, p.90). Although this thesis cannot be regarded as a legal study, in the second part of it we attempt content analysis of case law with a primary purpose to understand the reasoning of the European Court of Human Rights regarding hate speech and with a secondary purpose to compose an overview of recent practice of the Court regarding this issue.

In other words, content analysis conducted in the current thesis was based on two primary sources:

1. Selected case law of the European Court of Human Rights that was found in the online database of the court[4]. Selection of the judgments for the detailed analysis was based on time frame (2012-2019), as it is the same period when social media began to develop their own policies regarding hate speech, as well as on their significance and overall input to the understanding of the Court's approach towards hate speech. Significance was established by comparison of various secondary literature where these cases where

[4] The database of the European Court of Human Rights can be found here: http://hudoc.echr.coe.int/

repeatedly mentioned by legal scholars (among others: McGonagle, 2013; Bleich, 2013; Sottiaux, 2011; Tulkens, 2012; Buyse, 2014). .

As categorization is a central part of the qualitative content analysis (Mayring, 2000, p.3), an attempt to categorize the judgments was made based on the theoretical background discussed in paragraph 2.4.

2. Guidances, codes, community standards, official statements of the three social media platforms (Facebook, YouTube and Twitter) that were outlined and compared in the third part of the thesis. Additionally, online initiatives against hate speech were discussed in the paragraph 5.4

Observations regarding the comparison were presented in the Conclusion section.

4. European Court of Human Rights' approach towards hate speech

4.1. Definition and types of hate speech according to the European Court of Human Rights

Although the limits of freedom of expression were already discussed in the famous Handyside v. the United Kingdom case in 1976, the term "hate speech" was not included in the Convention and first appeared in four judgments of the Court in 1999, namely in Sürek v. Turkey (No. 1, par. 62); Sürek & Özdemir v. Turkey (par.63); Sürek v. Turkey (No. 4, par. 60) and Erdogdu & Ince v. Turkey (par. 54) (McGonagle, 2013, p.11).

Interestingly enough, in none of these four judgments does the Court provide a definition to hate speech. Instead, once again stressing responsibilities that come together with the right to freedom of expression, in three of four of these judgments the Court uses the same wording and warns media to be especially careful when publishing views of organizations that proclaim violence against the State as media should not become a mean "of dissemination of hate speech" (Erdogdu & Ince v. Turkey, 1999, par. 54).

However in the fourth of these cases – Sürek v. Turkey, 1999 – the Court went further and decided that whereas some information may lawfully "offend, shock or disturb", it the case Sürek v. Turke the commentary in question was "hate speech and glorification of violence" (see: Sürek v. Turke, 1999, par.62). Thus the Court established that there is a new type of speech – expression that goes beyond the protection given to the offensive

speech protected in Handyside v. the United Kingdom decision (McGonagle, 2013, p.12). It is also worth noticing that in all these four cases the Court connected hate speech with "glorification of violence".

Nevertheless, European Court of Human Rights did not provide a definition to hate speech – neither in 1999 not later. Without a definition, case law of the Court is the main source which is explored by scholars who attempt to analyse the Court's approach towards hate speech (McGonagle, 2013, p.12). While doing it, some of them underlined that it is important not to expand the term "hate speech" indefinitely (Tulkens, 2012, p.295).

It should be noticed that European Court of Human Rights itself distinguishes a much broader number of hate speech categories than was discussed in the theoretical part of this thesis. In 2019 the Court published a Factsheet on hate speech where a significant number of hate speech's types with examples of the respective most typical cases were listed (ECtHR, Factsheet hate speech, 2019, p.2-8):

- Ethnic hate (Pavel Ivanov v. Russia, 2007);
- Incitement to violence, support of terrorist activity (Roj TV A/S v. Denmark, 2018);
- Negationism and revisionism (mentioned earlier Garaudy v. France, 2003, also M'Bala M'Bala v. France, 2015);
- Racial hate (Glimmerveen and Haqenbeek v. the Netherlands, 1979);
- Religious hate speech (Norwood v. the United Kingdom, 2004);

- Speech that constitutes “Threat to the democratic order” (among others - Schimanek v. Austria, 2000);

- Speech that constitutes “Apology of violence and incitement to hostility” (discussed above Sürek (no.1) v. Turkey, 1999, also Gündüz v. Turkey, 2003);

- Hate speech in “Circulating homophobic leaflets” (mentioned above Vejdeland v. Others v. Sweden, 2012);

- Speech that condones terrorism (Leroy v. France, 2008, also Stomakhin v. Russia, 2018);

- Speech that condones war crimes (Lehideux and Isorni v. France, 1998);

- Speech that denigrates national identity (Dink v. Turkey, 2010);

- Speech that amounts to extremism (Ibragim Ibragimov and Others v. Russia, 2018);

- Displaying symbolic with “controversial historical connotation” (Fáber v. Hungary, 2012);

- Speech that incites ethnic hatred (Balsytė-Lideikienė v. Lithuania, 2008);

- Speech that incites national hatred (Hösl-Daum and Others v. Poland, 2014);

- Speech that incites racial discrimination or hatred (among others: Féret v. Belgium, 2009, Le Pen v. France, 2010, Perinçek v. Switzerland, 2015, Šimunić v. Croatia, 2019);

- Speech that incites religious intolerance (Erbakan v. Turkey, 2006);

- Speech that insults states officials (Stern Taulats and Roura Capellera v. Spain, 2018);

- Hate speech that is primary disseminated in the Internet (among others: Delfi AS v. Estonia, 2015, Pihl v. Sweden, 2017, Nix v. Germany, 2018, etc.).

Whereas this typology seems to be overwhelmingly detailed and not completely clear (for instance, the difference between ethnic hate and speech that incites ethnic hatred seems to be very blurred, as well as the difference between religious hate speech and speech that incites religious intolerance), it provides a full typological overview of the cases of the European Court of Human Rights that concern with information that the Court itself qualifies as hate speech. One could argue that it underlines the idea that it is not possible to provide a single finite definition to hate speech which instead should be better regarded as an open spectrum of speeches.

4.2. ECtHR's algorithm of establishing hate speech

European Court of Human Rights has repeatedly reiterated that there is an obligation to draw a line between expressing views (including such that can "offend, shock or disturb") and distributing information that represents incitement to extremism or any other form of hate speech (ECtHR, Factsheet hate speech, 2019, p.1). The Court proposes achieving this by applying two principal legal norms: Articles 10 and 17 of the European Convention on Human Rights. Article 17 – "Prohibition of abuse of rights" is to be applied when commentary in question attempt to destroy the fundamental values of the Convention, Article 10 – in all the other cases (ECtHR, Factsheet hate speech, 2019, p.1). For example, the limitations of the Article 10 must be applied when the commentary in question threatens public safety, health or the reputation or rights of others, etc. (ibid).

Nevertheless, this algorithm remains too vague to solve the collision of freedom of expression and hate speech in practice. Quite often it is not so easy to determine, for instance, whether a message represents someone's views on national security or territorial integrity or contains approval of terrorist's attacks. Is an article an example of journalist's provocation or the violation of the rights of others, do we need to protect moral in this or that case or do we need to protect the information, which shocks and disturbs, etc.

As was briefly mentioned in paragraph 2.1 of this thesis, the Court itself developed a classic test that is supposed to determine whether Article 10 was violated and that is codified in Article 10.2 of the Convention. Firstly, any interference with freedom of expression must be prescribed by law (European Convention on Human Rights, 1950, Art.10.2). Secondly, any restrictions of freedom of expression must have a "legitimate aim" – among others, prevent disorder and crimes or protect the reputation or rights of other people. Thirdly, these restrictions must be "necessary in a democratic society" (ibid). As emphasized by McGonagle, within these rules states-members of the Council of Europe may regulate freedom of expression themselves which explains why there is no consensus among different jurisdictions regarding what content, how and how much must be regulated (McGonagle, 2013, p.9).

It must be emphasized that the European Court of Human Rights does not replace national authorities in its decisions, but rather reviews their judgments for correspondence with the Convention's values (ibid). In other words, the Court's mission is to review national interpretation of the Convention (ibid).

4.3. Overview of ECtHR's hate speech practice in 1976-2012

Whereas the scope of this thesis does not allow us to conduct a thorough full-scale study of all the judgments related to hate speech regulation in the 60-year old history of the European Court of Human Rights, analysis of the recent case law requires at least a brief overview of selected historical and most influential judgments that formed modern Court's approach towards hate speech regulation. The general line is that the Court seeks harmonization of freedom of expression with protection of other human rights which involves protection from hate propaganda (Sottiaux, 2011, p.42).

The famous case Handyside v. United Kingdom, discussed above, took place in 1976 and is usually debated and quoted in any scholarly publication as an earliest example of the Court's approach towards hate speech regulation. In fact, the Court itself quotes Handyside v. United Kingdom in its Factsheet and in countless hate speech judgments (ECtHR, Factsheet hate speech, 2019, p.1). Still in the early period of its existence the Court ruled in the case Glimmerveen and Hagenbeek v. the Netherlands, 1979, that distribution of leaflets by far right political leaders in which they appealed to "our white people" and claimed to remove all foreign workers, constituted a form of racial discrimination (Glimmerveen and Hagenbeek v. the Netherlands, 1979, par.4). Bleich noted that Glimmerveen and Hagenbeek asserted that their leaflets were protected by freedom of speech provisions of the Convention, but the Court quoted Article 17 discussed above and dismissed their claims (2013, p.292).

Another case that happened before the Court introduced the term "hate speech" in 1999, was Jersild v. Denmark, 1994, in which a Danish journalist Jersild was convicted for spreading racist statements in the interview with right-wing group that he had conducted. This case is significant because it established a need for a journalistic responsibility when dealing with extremists views: the journalist in this case was convicted mainly because he failed to contradict his interviewees or distance himself from their racist statements (McGonagle, 2013, p.13).

The analysis of the scholarly literature indicates that the number of hate speech cases appears to increase in the 2000th. After the four judgments in 1999 that introduced the term "hate speech" in connection with "glorification of violence" (see par. 4.1), a series of important rulings took place one by one. The judgment in Garaudy v. France, 2003, established Holocaust denial as unlawful and incitement to hatred against Jews, although in one of the previous cases the Court ruled that generally Article 10 of the Convention protects debate about unsettled historical events and public figures about whom different opinions may exist (Lehideux and Isorni v. France, 1998, par. 45).

The case Gündüz v. Turkey, 2003, introduced debate about religious hate speech in the practice of the European Court of Human Rights. In this case the Court overturned Turkish courts' decision that discussions on Islam and secularism constituted incitement to hatred based on religion. Instead, European Court of Human Rights ruled that defending religious beliefs without incitement to violence does not amount to hate speech

(Gündüz v. Turkey, 2003, par.51). In other words, the Court underlined incitement to violence as a necessary component of hate speech.

The Court reiterated its position a year later in the case Norwood v. United Kingdom, 2004, which concerned a member of a British far right party who displayed out of a window of his apartment a poster with Twin Towers if flames, accompanying by the call to eliminate Islam from Britain in order to protect its people (Norwood v. United Kingdom, 2004, par. 2). The Court decided that this fact amounted to a public attack on Muslims in the United Kingdom and so a conviction of Norwood by British courts did not constitute a breach of the Article 10 (ibid, par.4). It is worth noticing that the term "hate speech" did not appear in this decision even five years after its introduction in the Court's practice.

According to Tulkens, the essence of today's position of the Court was expressed in the case Erbakan v. Turkey, 2006 (2012, p.279). In this case the Court evaluated decisions of Turkish courts that convicted a future Turkish Prime Minister Erbakan for inciting hatred in his political speech in which he discussed differences between religions and races. The European Court of Human Rights concluded that, as speech in question did not create imminent danger, the penalty imposed by the national courts would have negative consequences for the political debate in the country. Thus, the Court underlined negative consequences as important factor in establishing hate speech. Simultaneously, the Court reiterated that in democratic society it may be necessary to prevent dissemination of

expressions that incite or justify hatred on the grounds of intolerance (Erbakan v. Turkey, par. 56).

Another significant case related to ethnic hate speech is Pavel Ivanov v. Russia, 2007, in which the Court convicted Russian journalist Ivanov in spreading hatred via his newspaper. In this case the speech in question was series of anti-Semitic publications in which Ivanov accused the entire ethnic group of Jews in all Russian problems, and called for their exclusion from social life (Pavel Ivanov v. Russia, 2007, The Facts). The Court had no doubts that this sort of speech constituted a violation of the underlying values of the Convention (ibid, The Law, par.1). However, in this judgment the Court did not use the term "hate speech" but referred to Ivanov's publications as "incitement of ethnic, racial and religious hatred" (ibid, the Facts). This underlines the fact that the use of the term "hate speech" is not unanimous, and in 2007 the judges still preferred to use more general terminology.

Incitement to racial hatred was further explored in the judgements Soulas and Others v. France, 2008 and Féret v. Belgium, 2009. In both these cases the speech – a published book and leaflets respectively – was directed against immigrants and incited hatred and violence against them. In the first case, the author of the book in question went so far as to propagate an "ethnic war" (Soulas and Others v. France, The Facts). In the second, a chairman of a political party Féret called to stop "Islamification of Belgium" (Féret v. Belgium, 2009, par.9). In both cases the European Court of Human Rights supported the

national courts and reiterated that conviction of the applicants was in the interest of order in the society.

Finally, two more case that must be mentioned in this very brief overview of the European Court of Human Rights' practice in the period 1976-2012 and related to hate speech, are the cases Le Pen v. France, 2010 and Dink v. Turkey, 2010. The first of them is usually quoted together with the Féret v. Belgium, 2009, because of the similar nature. Thus, the applicant in this case was a prominent French politician Le Pen whom French courts found guilty in incitement of hatred and discrimination against Muslims in France for his far-right statements in an interview. The European Court of Human Rights upheld the decisions of the national courts, confirming that political comments that presented a whole religious community as a national threat are likely to result in "rejection and hostility" against these people (Le Pen v. France, 2010, The Law, par.1).

On the contrary, the case Dink v. Turkey, 2010 represents a situation where the state – Turkey – condemned a journalist Firat Dink for his articles on the identity of Turkish people of Armenian origin. A year later, the journalist was killed. The relatives brought the case to the European Court which concluded that there was no "pressing social need" to find the journalist guilty by national courts because the articles did not incite violence or revolt and were not disrespectful or insulting (Dink v. Turkey, 2010, par.136). The journalist was merely expressing his opinions on the issue of public interest (ibid, par.135). It must be noted that additionally the Court highlighted the fact that authorities failed to protect the journalist (ibid, par.91).

To sum it up, the argumentation of the Court in the discussed judgments is presented in the following Table 1:

Table 1. Argumentation of the ECtHR in selected hate speech-related cases 1976-2012

Judgement	Argumentation in support of the utterance in question **being hate speech**	Argumentation in support of the utterance in question being protected speech (**not hate speech)**
1. Handyside v. United Kingdom, 1976		Ideas that can "offend, shock or disturb" must be free in the interests of "pluralism, tolerance and broadmindedness" (par.49)
2. Glimmerveen and Hagenbeek v. the Netherlands, 1979	Political leaflets with appeal to remove all foreign workers, constituted a form of racial discrimination (par.4)	
3. Jersild v. Denmark, 1994	Journalist convicted because he failed to contradict his interviewees or distance himself from their racist statements	
4. Lehideux and Isorni v. France, 1988		Generally, the Convention protects debate about unsettled historical events and public figures about whom different opinions may exist (par.45)
5-8. Sürek & Özdemir v. Turkey, 1999, Sürek v. Turkey, 1999 (No. 4), Erdogdu & Ince v. Turkey, 1999 Sürek v. Turkey, 1999	First use of the term "hate speech" by ECtHR, no definition is given, but responsibilities are highlighted: media should not become a mean "of dissemination of hate speech" (Erdogdu & Ince v. Turkey, 1999, par. 54). The Court established a connection between hate speech and "glorification of violence" (Sürek v. Turke, 1999, par.62)	

9. Garaudy v. France, 2003	Holocaust denial is unlawful and constitutes incitement to hatred against Jews (par.45)	
9. Gündüz v. Turkey, 2003		Defending religious believes without incitement to violence does not amount to hate speech (par.51)
10. Norwood v. United Kingdom, 2004	Call to eliminate Islam from Britain amounted to a public attack on Muslims in the United Kingdom (par.4)	
11. Erbakan v. Turkey, 2006		Because speech in question did not create imminent danger, the penalty imposed by the national courts would have negative consequences for the political debate (par.56)
12. Pavel Ivanov v. Russia, 2007	Accusation of the entire ethnic group and call for Jews to be excluded from social life constituted a violation of the underlying values of the Convention and incitement of ethnic hatred (The Law, par.1)	
13. Soulas and Others v. France, 2008	Propaganda of an "ethnic war" is incitement to hatred	
14. Féret v. Belgium, 2009	Distributing leaflets with appeal to stop "Islamification of Belgium" is incitement to hatred (par.9)	
15. Le Pen v. France, 2010.	Political comments that present a whole religious community as a national threat are likely to result in "rejection and hostility" against these people and therefore, conviction of the speaker is relevant and sufficient (The Law, par.1)	
16. Dink v. Turkey, 2010		There was no "pressing social need" to find the journalist guilty by national courts because the articles did not incite violence or revolt, were not disrespectful or insulting and the journalist was

		expressing his opinions on the issue of public interest (par.135-136)

In the next paragraph more detailed analysis of the arguments provided in the judgments passed by the European Court of Human Rights in 2012-2019 – already after the widespread of social media – will be provided.

4.4. ECtHR's case law concerning hate speech in 2012-2019

Despite the typology found in the Factsheet hate speech of the European Court of Human Rights and discussed above, we propose to categories the analysis by content of the cases, as it was done by Tulkens who distinguished only four categories of hate speech: glorification of violence, religious intolerance, racial hate speech and sexual orientation hate speech (2012).

In this paragraph following fifteen judgments of the European Court of Human Rights from the period 2012-2019 will be analyzed in details:

- Political/racial//ethnic hate speech (Aksu v. Turkey, 2012; Öner and Türk v. Turkey, 2015; Balázs v. Hungary *2015;* Perinçek v. Switzerland, 2015; Király and Dömötör v. Hungary, 2017);

- Religious hate speech (Raelien Suisse v Switzerland, 2012; Belkacem v. Belgium, 2017);

- Sexual orientation hate speech (Vejdeland and Others v. Sweden, 2012; Identoba and Others v. Georgia, 2015; M.C. and A.C. v. Romania, 2016);

- Online hate speech (Delfi AS v. Estonia, 2015; Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016; Pihl v. Sweden, 2017; Smajić v. Bosnia and Herzegovina, 2018; Savva Terentyev v. Russia, 2018).

The results of this analysis will be presented in Table 2 in paragraph 4.5.

4.4.1. Political, racial, ethnic hate speech

Aksu v. Turkey, 2012

As was mentioned in the chapter 4.1 of the present paper, the European Court of Human Rights distinguishes, among others, political, racial, national and anti-constitutional hate speech. However, as will be evident from the detailed analysis of the judgments, this separation is not always clear and in some cases, is very blurred. That is why, for the purpose of the present thesis, it was decided to combine the next five judgments in one category, namely, political, racial and ethnic hate speech.

The first of the five judgments to be discussed in this category is the case of Aksu v. Turkey, 2012 where the alleged hate speech was contained in the book "The Gypsies of Turkey" written by the Associate Professor Ali Rafet Özkan and published by the Turkish Ministry of Culture. The book in question was a research about people of Roma origin in which the author describes the ethnic and cultural features of the respective nation. Among other things, the author claimed that people of Roma origin "were engaged in illegal activities" and lived as "thieves, pickpockets, swindlers, robbers, usurers, beggars, drug dealers, prostitutes and brothel keepers" (Aksu v. Turkey, 2012, par.14). These particular remarks attracted the attention of the applicant of Roma origin and he filled the complaint in 2002, claiming that the passages were insulting and humiliating. The national courts in long and complicated proceedings found no proof that such remarks constitute an attack on the applicant or the other people of Roma origin.

In a separate but related case, the same applicant claimed that several entries in the new published Turkish dictionary contained discriminatory and insulting information about the Roma people. For instance, the applicant especially mentioned such entries as "Gypsy tent': a dirty and poor place" and 'Becoming a Gypsy': displaying miserly behavior" (ibid, par.28). Several national courts dismissed the complaint on the ground that the entries reflect historical and sociological reality and were not meant to insult a particular ethnic group.

The case came to the European Court of Human Rights that investigated the complaint and found a need for balance between the applicant's rights and public interest in

freedom of expression. Restating, that freedom of expression is one of the most essential foundations of a democracy, the Court reminded about the information that can "offend, shock or disturb" and that was protected in the case *Handyside v. the United Kingdom, 1976*. The Court then proceeded with the reminder that freedom of expression may be subject to the exceptions described in the second part of the Article 10 of the European Convention on Human Rights, but the necessity of such restrictions must be established "convincingly" (ibid, par.64).

Once again, the Court underlined that the analysis of the passages from the book in question must be conducted in the context of the whole book, including the research method used by the author who contacted the members of the Roma community, as well as the police and relevant authorities (ibid, par.20). Furthermore, the author claimed that he lived with the Roma community and observed their lifestyle from the scientific point of view (ibid, par.11).

Additionally, the European Court stated that the applicant had an opportunity to review his case before two different levels of national jurisdiction and the Ministry of Culture windrowed the remaining copies of the book. Taking it all into account, the Court concluded that the necessary balance between the rights of the Roma community and freedom of expression was found, even considering the vulnerable position of the Roma people in the country. According to the court, the government should do more to battle the negative stereotypes about the Roma community, but in the present case the Turkish authorities did not misuse their power and did not violate anybody's rights. The

expressions used in the book are part of a spoken daily Turkish, as much as the dictionary's entries from the second part of the case. The entries must have been labeled as "insulting", but an absence of this note was not enough to constitute a violation.

It is necessary to emphasize that once again the court considered the context around particular words and phrases, and mentioned both the case of Handyside v. the United Kingdom, 1976 and the European Convention on Human Rights while establishing a balance between offensive speech and freedom of expression.

Öner and Türk v. Turkey, 2015

Another case that was considered to be an example of alleged political and racial hate speech also originated in Turkey. The applicants in the Öner and Türk v. Turkey, 2015, attended public celebrations and made speeches about Kurdish people. Among others, the applicants talked about the civilians killed by the security forces in Cizre, Nusaybin and Şırnak, the poisoning of the Kurdish leader and claimed that the authorities did not supported democratization and did not solve the Kurdish problem (Öner and Türk v. Turkey, 2015, par.6).

Shortly after the event, the public prosecutor charged the applicants with terrorist propaganda; the national court found them guilty and sentenced them to one year and eight month of imprisonment. Though the sentence was eventually suspended, the

applicants filed a complaint to the European Court of Human Rights, claiming that their right for freedom of expression guaranteed to them by Article 10 of the Convention was violated. The applicants emphasized that in their speech they spoke about the necessity to find peaceful methods to solve the Kurdish problem, did not advocate any kind of violence and did not called upon people to commit any kind of illegal actions (ibid, par.19).

After reviewing the case, the European Court concluded that there was an interference with the applicants' rights to freedom of expression and it was not necessary in a democratic society. From the Court's perspective, the speech in question constituted a critical assessment of the authorities and did not contained terrorist propaganda. Moreover, the Court specifically mentioned that the applicant's speech did not "encourage violence, armed resistance or an uprising" (ibid, par.24) and did not incited violence by "by instilling a deep-seated and irrational hatred against identifiable persons" – and thus was not hate speech (ibid).

Interestingly enough, this judgment is one of the very few instances where the European Court of Human Rights specifically mentioned the term "hate speech", adding to it a significant characteristic. Thus, from this judgment it is possible to assume that the Court understands under hate speech a speech that is capable of inciting violence by appealing to the irrational hatred against a person or a group of people.

Additionally, the case Öner and Türk v. Turkey, 2015 is a good example of blurring of the categories of political and ethnic hate speech or alleged hate speech as the initial speech of the applicants was about the Kurdish problem and their discrimination in the society, as well as about the authorities' policy towards this issue.

Balázs v. Hungary 2015

The case Balázs v. Hungary 2015, can also be classified as a political or ethnic hate speech case. In it, the applicant of Roma origin and his girlfriend were leaving a club when three men approached and started to insult them with "Dirty gypsy, do you need a cigarette? Here is money!" (par.10). The situation developed in a fight that was concluded by the police arrival. Subsequently, the national courts failed to establish a link between the insult with national hatred and the fight, and the case came into the European Court.

The Court concluded that the state has an obligation to establish any possible racist motive and ethnic hatred behind the attack. Even though proving racial motivation could be difficult, the national courts should have taken into account that people of Roma origin are especially vulnerable in the Hungarian society and require an additional protection (ibid, par.53).

Moreover, the Court underlined that any racist verbal abuse is highly relevant in cases involving ethnic minorities (ibid, par.61). When only the victim's characteristic is

involved, the attack can be classified as hate crimes even if the motives were mixed (ibid, par.70). Therefore, concludes the court, the state clearly violated the rights of the applicant by not conducting a thorough investigation (ibid, par.75).

While relevant for the present paper, it is necessary to mention that in the case Balázs v. Hungary 2015 the term "hate speech" was not used as such in the court's assessment part of the judgment, but was substituted by the phrase "racist verbal abuse" which in the present case can be regarded as a synonym. Regarding the typology, the case Balázs v. Hungary 2015 provides an example of ethnic hate speech.

Perinçek v. Switzerland, 2015

Finally, the last selected case found for the category of political, racial and ethnic hate speech is the case Perinçek v. Switzerland, 2015. In this case the applicant was a doctor of law and a chairman of the Turkish Workers' Party who in 2005 took part in a various press conferences where he publicly denied the genocide of the Armenian people by the Ottoman Empire in 1915. In one of his speeches, he claimed that "the allegations of the 'Armenian genocide' are an international lie" and that the Kurdish and the Armenian problem "did not even exist" (Perinçek v. Switzerland, 2015, par.13).

After the conferences the Switzerland-Armenia Association filed a complaint against the applicant and the Lausanne District Police Court found the applicant guilty of racial

discrimination as the Armenian genocide was a proven fact (ibid, par.186). After the appeal in the national courts, the case came to the European Court of Human Rights where the applicant claimed that the courts breached his freedom of expression.

In 2013 a Chamber of the Court found a violation of Article 10 of the Convention, but at the Government's request the case came to the Grand Chamber of the Court which analyzed, among others, whether the speech in question can be regarded as hate speech by comparing it to the similar cases such as, among others, Gündüz v. Turkey, 2003 and Erbakan v. Turkey, 2006.

After this analysis, the Court concluded that the national courts failed to establish a social need or necessity of the applicant's conviction in a democratic society, as the authorities overstepped their power and suppressed a debate with huge public interest (ibid, par. 241). The mere denial of Armenian genocide is insufficient to be qualified as an incitement of hatred towards a particular nation and, accordingly, the applicant had the right for open discussion of the issue even if it is sensitive and may be offensive, as it is a fundamental aspect of freedom of expression (ibid, The Court's assesment).

Once again it is necessary to emphasize that the Court did not only analyzed the particular words or speech, but regarded it in a broader context of the historical and political situation. Moreover, the Court focused on whether the speech in question had a public interest and whether a conviction of the speaker was necessary in a democratic society – two more crucial questions that can be used in establishing whether particular

content constitutes hate speech or not. From this perspective, the difference between the case Király and Dömötör v. Hungary, 2017 analyzed below and the cases Perinçek v. Switzerland, 2015 and Öner and Türk v. Turkey, 2015 is especially evident.

Király and Dömötör v. Hungary, 2017

Finally, the fifth case to be analyzed in this category is Király and Dömötör v. Hungary that was passed in January 2017. In this case the applicants – two Hungarians of Roma origin – complained that the state failed to protect them during the anti-Roma demonstration that has taken place in 2012. This demonstration was organized by the far-right groups who described it on their websites as "against Roma criminality" (Király and Dömötör v. Hungary, 2017, par.7*).*

During the events the leaders of the far-rights groups called Roma "not "normal", and claimed that Roma criminality was "omnipresent" in Hungary and that this ethnic group brought "only destruction, devastation and fear" (ibid, par.10). On this ground, the leaders called the participants to "sweep out the "rubbish" from the country" and stated that the Hungarians had rights to use all possible means to achieve this purpose (ibid).

The applicants who were in the neighborhood during the demonstration heard these and other speeches and later submitted a complaint to the national courts. According to the applicants, the police and eventually the state failed to protect them from the racist

threats. The Hungarian courts, however, concluded that the police acted professionally and had no legal basis to stop the demonstration (ibid, par.15).

The European Court of Human Rights did not question the actions of the Hungarian police, but concentrated mainly on the intimidating speeches that the applicants were exposed to due to their ethnicity. While the Hungarian courts did not investigate this part on the ground that the statements made during the march were not included in the initial complaint, the European Court of Human Rights reminded about the discrimination of Roma minority in the country and repeated examples of hate speech and even hate-motivated killings (ibid, par.40).

According to the European Court of Human Rights, harassment motivated by racism with no physical violence was still a crime and a violation of the applicants' rights. The Court especially underlined a tense political situation in the country (ibid, par.73) and acknowledged that in all similar cases the outcome depended on the various factors as the Court's approach to the cases with an alleged hate speech was "highly context-specific" (ibid. par.74).

In other words, the Court did not analyze particular phrases or words that were said during the far-right demonstration, but rather the whole situation of hatred and threat that was created. The context appears to be more significant for the Court than the precise utterances. In the case Király and Dömötör v. Hungary, 2017 the European Court concluded that national courts failed to investigate the context of the speeches and, by

establishing that the statements were hateful and abusive, but did not exited any violence, eventually failed to protect the rights of the applicants.

4.4.2. Religious hate speech

Raelien Suisse v Switzerland, 2012

Two of the selected recent judgments of the European Court of Human Rights could be classified as religious hate speech: Raelien Suisse v Switzerland, 2012 and Belkacem v. Belgium, 2017. In the first of these cases the applicant was an association of the Raelian Movement that was registered in 1977 with an aim to make first contacts with extraterrestrials and establishing good contacts with them (Raelien Suisse v Switzerland, 2012, par.10). In 2001 the organization requested the authorization from the police to conduct a poster campaign during which a poster "The Message from Extraterrestrials" was supposed to be demonstrated (ibid, par.14). It was expected that together with the contact details of the organization the phrase "Science at last replaces religion" would have been written on the poster (ibid*).*

The authorities, however, refused to grant permission to the event, citing the previous national courts' decisions that the movement was engaged in the activities against the public order (ibid, par.15). The national court once again stated that the Raelian Movement could not have a religious freedom because it was regarded as "a dangerous

sect" as it has advocated, among other things, for the human cloning and "geniocracy" (ibid, par, 16).

After national courts in Switzerland upholded this decision, the case went to the European Court that reiterated the importance of establishing the nature of the speech (ibid, par.61). It was clear that the poster wanted to draw attention of the public to the movement with a religious connotation and did not contribute to the political debate in Switzerland, thus being closer to the commercial advertisement than to the political speech, which gives authorities more freedom (ibid, par.62). After establishing that fact, the Court underlined that different states may have reasons to impose restrictions in such issues (ibid, par.63). In the present case the Court found a prohibition of the poster demonstration to be a reasonable restriction in order to protect health, morals and the rights of others in the society (ibid, par.72).

Moreover, the Court found the position of the authorities to be in balance, as they did not banned the whole association or its website and even without the poster campaign it had other means to disseminate its ideas and express its beliefs (par.73). Accordingly, the decision of the police to deny the permission was relevant and proportional and answered the need of the society and, therefore, there was no violation of Article 10 (par.76).

It is important to emphasize that the official judgment of the Court was complemented by the dissenting opinion of one of the judges who wrote, among others, that the speech in question – in particular the phrase "Science at last replaces religion" did not constitute

religious hate speech and did not denigrate any religion (ibid, Scientific atheism). The judge then mentioned the Article 10 of the Convention and his belief that it prohibits the state to play a role of a watchman and to prescribe what is right or wrong (ibid). Accordingly, the judge concluded that ban of the poster did not correspond to the social need in his opinion (ibid).

While analyzing this case, it is necessary to notice that the Court once again examined the speech in a broader context, namely, the nature of the organization and its beliefs, as well as the purpose of the speech in question. That is similar to the Court's approach in the cases with alleged political and national hate speech where it also took into account whether the speech was made with a purpose to incite hatred and intolerance, or in order to trigger public debate.

Belkacem v. Belgium, 2017

Another selected case – Belkacem v. Belgium, 2017 – represents a more typical example of religious hate speech than the previous case. In this case the applicant was convicted by Belgian courts for incitement to hatred in his YouTube videos where he called on his audience, among other things, to fight with non-Muslims and "teach them a lesson" (Belkacem v. Belgium, 2017, par.8). In the European Court the applicant claimed to merely express his opinion and exercise his freedom of expression.

The Court, however, unanimously and "without any doubts" concluded that the applicant disseminated hateful content and sought to incite violence against all non-Muslims which is clearly against the values of the Convention (ibid, par.33). Additionally, the Court highlighted that defending Sharia while advocating for violence can amount to hate speech and that states have an obligation to contradict "religious fundamentalism" (ibid, par.34).

4.4.3. Hate speech based on sexual orientation

Vejdeland and Others v. Sweden, 2012

As was briefly mentioned in the theoretical part of this thesis, the case Vejdeland and Others v. Sweden, 2012 was the case where the European Court of Human Rights recognised homophobic hate speech for the first time (McGonagle, 2013, p.12). The applicants in this case distributed leaflets in a school in which they claimed that homosexuality was a "deviant sexual proclivity" responsible for HIV and AIDS and destructive for a society (Vejdeland and Others v. Sweden, 2012, par.8). In their defence the applicants claimed that they merely wanted to start a debate (ibid, par.10).

The European Court, however, concluded that even of these statements did not propagate hateful acts as such, they were still serious allegations that reinforced prejudices (ibid, par.54). The Court reiterated that discrimination based on sexual orientation should be

considered as serious as discrimination on the ground of race, sex, colour or origin (ibid, par.42). Therefore, national courts did not violate Article 10 by convincing the applicant.

Identoba and Others v. Georgia, 2015

Another case of the European Court that concerns sexual orientation hate speech is Identoba and Others v. Georgia, 2015. In this case the complaint was filed by 14 people and a non-governmental organization set up to promote and protect the rights of LGBT people. According to the applicants, in May 2012 during a peaceful demonstration for the International day against Homophobia they were insulted and threatened by the counter-demonstration of members of a religious group (Identoba and Others v. Georgia, 2015, par.10). In particular, the applicants were accused of being "seek", "immoral" and "perverts" and on this ground the participants of the religious demonstration stated that they "should be burnt to death" and "crushed" (ibid, par.13 and par.15). The threats were followed by the attacks and some of the applicants suffered physical trauma (ibid, par.18)

National authorities failed to investigate the case properly and the case came to the European Court which concluded that the applicants clearly became the target of hate speech and aggressive actions (ibid, par.70). The whole situation contained an intense anxiety, with a homophobic bias as an additional factor (ibid).

The domestic authorities, however, failed to investigate the homophobic motives of the event and did not determine homophobic hate speech (par.77). In its conclusion, the Court once again stated that the applicants were left in anxiety and fear by violence which "consisted mostly of hate speech and serious threats" (ibid, par.70) and thus there was a clear violation of the applicants' rights.

M.C. and A.C. v. Romania, 2016

In a similar case - M.C. and A.C. v. Romania, 2016 - hate speech again was addressed towards the participants of the LGBT-rally, but this time in Bucharest. After the event, the applicants were physically attacked by a group of six young men and a woman who were shouting "You poofs go to the Netherlands!" (M.C. and A.C. v. Romania, 2016, par.9).

Similar to the previous case, Romanian authorities failed to protect the rights of the applicants. The European court of human rights found that the authorities did not investigate homophobic motives behind the attack especially giving that the speech in question was "clearly homophobic" hate speech (ibid, par.124).

In other words, in both Identoba and Others v. Georgia, 2015 and M.C. and A.C. v. Romania, 2016 the European Court had little doubt about the nature of the speech and concluded it to be sexual orientation hate speech on the ground of the context, on one side, and taking into account the whole situation and the following attacks, on the other.

4.4.4. Online hate speech cases in the practice of the ECtHR

Delfi AS v. Estonia, 2015

Five of the selected cases could be classified as online hate speech. Chronologically first of them – Delfi AS v. Estonia, 2015. The applicant of this case was a company that owned the largest news portal in Estonia which publishes up to 330 news per day both in Estonian and Russian (Delfi AS v. Estonia, 2015, par.11). In 2006 one of the articles published by the news portal attracted the attention of the readers and consequently received 185 comments (ibid, par.17). Approximately twenty of them contained personal threats and offensive language against a person mentioned in the article (ibid).

For instance, among the comments were “bloody shitheads”, “go and drown yourself”, “a good man lives a long time, a shitty man a day or two”, etc. (ibid, par.18). The lawyers of the concerned person filled a lawsuit and at the same day the news portal deleted the offensive comments (ibid, par.19). The national court, however, decided that as the news portal reserved the right to delete inappropriate comments, it has taken insufficient measures to protect the rights of others (ibid, par.26). Moreover, the national court of appeal reiterated that the news portal was not a technical intermediary, but a “provider of content services”and therefore had respective obligations (ibid, par.29).

Disagreed with the decisions of the national juridical system, the Delfi Company brought the case to the European Court of Human Rights that eventually shared the position of the national courts. Thus, the European Court emphasized that, whereas it acknowledges how beneficial Internet can be for freedom of expression, it must be taken into account that all kinds of unlawful speech, including hate speech, can be easily disseminated online (ibid, par.110). Therefore, it is important to find a balance between freedom of expression, on one the hand, and prohibiting unlawful speech that constitutes a violation of personality rights and other underlying values of the Convention (ibid).

According to the Court, Internet news portals have "duties and responsibilities" when their users disseminate unlawful speech through their sites (ibid, par.113). Taking into account that the applicant owned one of the biggest news portals in Estonia, the nature of some comments on the portal was known before the present case (ibid, par.117) As was established by the Supreme Court of Estonia and reiterated by the European court, the comments in question constitute hate speech and speech that "directly advocated violence" (ibid). That is why the automatic filter-system used by the portal to filter the instances of hate speech was insufficient and failed to prevent the violation of the rights of others (ibid, par.156).

Based on these considerations and taking into account the nature of the comments, the Court upheld the decision of the national courts and concluded that there was no violation of Article 10 of the Convention in convicting the news portal.

Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016

The second case concerned with online hate speech is Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016. The case involves a self-regulatory body of Hungarian Internet content providers and a major news portal of Hungary who were the applicants of the case.

In 2010, when both applicants allowed users' comments of their portals without premoderation, the applicants published an article that illustrated their opinion about two real estate websites (Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016, par.11). The article attracted the attention of the audience and among the comments were phrases like "They have talked about these two rubbish real estate websites a thousand times already" and "Is this not that Benkő-Sándor-sort-of sly, rubbish, mug company again?" (ibid, par.12).

The company-owner of the websites filled a lawsuit, accusing the applicants that their opinion and the users' comments were false and had a negative influence on their reputation (ibid, par.15). The applicants removed users' comments, but the national courts still found them guilty in infringing the websites' company-owner' reputation (ibid, par.17).

When the case came to the European Court of Human Rights, it compared it to the case Delfi AS v. Estonia, 2015. However, the court concluded the Hungarian case was

different: after the analysis of the speech, the court found that the users' comments were "offensive and vulgar", but not unlawful and did not amount to hate speech (ibid, par.64).

Once again, the court underlined the necessity to find a balance between the rights of all concerned parties. If the users' comments constitute hate speech, the state may impose liability upon news portals. But the present case did not involve such context and therefore the court concluded that there had been a violation of the rights of the applicants (ibid, par.91).

Pihl v. Sweden, 2017

In a similar case, the applicant was a subject of an online defamation, namely, of an accusation of being a member if a Nazi party (Pihl v. Sweden, 2017, par.3). The applicant replied online that this information was wrong and the administration of the blog in question deleted the previous entry (ibid, par.6). Still, the applicant claimed that national courts refused to hold the administration of the web-site liable and did not protect his reputation (ibid, par.15).

The European Court, however, declared that in this case a balance between freedom of expression and applicant's rights was found by national courts as the defamatory post was quickly removed and an apology issued (ibid, par.29 and 30). Additionally, the Court

highlighted that although offensive, the post did not incite violence and did not amount to hate speech against the applicant (ibid, par.37).

Smajić v. Bosnia and Herzegovina, 2018

The applicant of this case was convicted for posting online statements in which he described military actions that might be undertaken by Bosnians against Serb villages in the event of a new war and expressed insults against Serbs (Smajić v. Bosnia and Herzegovina, 2018, par.5). The applicant claimed to discuss a matter of public interest (ibid, par.29).

The European Court unanimously rejected these allegations, stating that conviction of the applicant was prescribed by law for the protection of the rights of others (ibid, par.32). Furthermore, the Court underlined that, while hypothetical, the applicant's statement related to a very sensitive subject in Bosnian society and so national courts had sufficient reasons to convict him (ibid, par.39).

Savva Terentyev v. Russia, 2018

The case Savva Terentyev v. Russia, 2018 can probably be regarded as a classic example of "Handyside v. United Kingdom rule" which states that certain expressions may

"offend, shock or disturb", but must, nevertheless, remain free in the interests of "pluralism, tolerance and broadmindedness" (Handyside v. United Kingdom, 1976, par.49). In this case an applicant posted remarks about Russian police, calling them "faithful dogs" and making other insulting remarks about them (Savva Terentyev v. Russia, 2018, par.10).

While acknowledging that the applicant's comments were expressed in an offensive form, the Court unanimously concluded that they did not incite hatred (ibid, par.84). The comments in question were "an emotional reaction" of the applicant to the abusive actions of the police and did not provoke any violence (ibid). Therefore, the Court stated the criminal conviction of the applicant for a sentence of one year's imprisonment was disproportionate and not justified by "pressing social need" (ibid, par.86).

4.5. Summary of the ECtHR's practice concerning hate speech

To summarize the analysis of the case law of the European Court, it is necessary to emphasize certain criteria that the Court employs in order to determine whether the speech in question is hate speech or not and what judgment to pass. Those criteria are:

1) Nature of the speech: whether it is capable of inciting violence or hatred or not;

2) Purpose of the speech: whether it is aiming at inciting hatred or not;

3) Significance of the speech: whether it has public interest and contributes to the public debate or not;

3) Context of the speech: whether there are tensions that accompany the speech in question or not;

4) Necessity of convincing the speech and the speaker in the democratic society;

5) Whether a balance between freedom of expression and rights of others was taken into account by the national courts, or not.

Further arguments of the European Court from the analyzed cases are summarized in Table 2 below.

Table 2. Argumentation of the ECtHR in selected hate speech-related cases 2012-2019

Judgement	Argumentation in support of the utterance in question **being hate speech**	Argumentation in support of the utterance in question being protected speech (**not hate speech)**
1. Aksu v. Turkey, 2012		A need for balance between the applicant's rights and public interest in freedom of expression is underlined. Necessity of restrictions of freedom of information must be established "convincingly" (par.64). The Court reiterates the importance of the context and concludes that balance between the rights of the Roma community and freedom of expression was found.
2. Öner and Türk		The Court mentions the term

v. Turkey, 2015		"hate speech", adding that it is capable of inciting violence by appealing to the irrational hatred against a person or a group of people (par.24). The Court concludes that the speech in question was not hate speech, but critical assessment of the authorities that did not contained terrorist propaganda and did not encourage violence (par.24)
3. Balázs v. Hungary *2015*	State has an obligation to establish any possible racist motive and ethnic hatred behind the attack (par.53). Any racist verbal abuse is highly relevant in cases involving ethnic minorities (par.61).	
4. Perinçek v. Switzerland, 2015		Public statements that concern a matter of public interest must have higher protection under Article 10 of the Convention par. (197). The conviction of the speaker was not necessary in a democratic society (par.158). The statement in question did not call for hatred or intolerance and the context lacked tensions, therefore the speech of the applicant must have been protected by the Article 10.

5. Király and Dömötör v. Hungary, 2017	The Court emphasized the importance of the context and states that cases with an alleged hate speech are "highly context-specific" (par.74).	
6. Raelien Suisse v Switzerland, 2012	It is important to determine the type of speech and its purposes. Speech with commercial purposes – such as commercial advertisement – should have less protection by the Article 10 than political debate (par.62). It is important to search for relevant and proportional restrictions to speech that raises concerns (par.75).In this case, the decision of the national courts to ban the poster was relevant (par.76).	
7. Belkacem v. Belgium, 2017	Advocating for religious fundamentalism while calling for violence can amount to hate speech (par.34). States have an obligation to prohibit it (ibid). General attack on people based on their religious beliefs or lack of such is incompatible with the values of the Convention and cannot be protected by the Article 10 (par.33).	
8. Vejdeland and Others v. Sweden, 2012	Even if certain statements do not propagate hateful acts as such but are serious allegations that reinforce prejudices, they might be prohibited (par.54 and 55). Discrimination based on sexual orientation should be considered as serious as discrimination on the ground of race, sex, colour or origin (par.42)	

9. Identoba and Others v. Georgia, 2015	Homophobic motives must be investigated in hate speech and subsequent threats. The context of the situation must be taken into account.	
10. M.C. and A.C. v. Romania, 2016	Authorities must investigate homophobic motives. The context and, if applicable, the following attacks must be taken into account.	
11. Delfi AS v. Estonia, 2015	The Court acknowledges that Internet can be beneficial for freedom of expression, while at the same time enabling dissemination of unlawful speech, including hate speech. Therefore, a balance between freedom of expression and prohibiting unlawful speech must be found. Internet news portals have "duties and responsibilities" when their users disseminate unlawful speech through their sites (ibid, par.113) In this case, automatic filter-system used by the portal to filter the instances of hate speech was insufficient and failed to prevent the violation of the rights of others (ibid, par.156).	
12. Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016		Analysis of the speech itself must be conducted. "Offensive and vulgar" comments do not necessarily amount to hate speech (par.64).

13. Pihl v. Sweden, 2017		Although offensive, the comment in question did not incite to violence and did not amount to hate speech (par.37). National court found a balance between freedom of expression and protection of the applicant's rights as the comment was quickly removed and an apology issued (par.29 and 30).
14. Smajić v. Bosnia and Herzegovina, 2018	Conviction of the applicant was prescribed by law for the protection of the rights of others (par.32). Even hypothetical statements might be prohibited, taking into account the context – in this case, sensitive subject of ethnic relations between Bosnians and Serbs (par.39).	
15. Savva Terentyev v. Russia, 2018		Insulting and offensive remarks might still be protected by the Article 10, provided that they do not incite hatred or violence (par.69). Criminal conviction is disproportionate for offensive speech without incitement to hatred (par.86).

5. Social media and online hate speech

As was illustrated by the judgments of the European Court of Human Rights, disputes related to the alleged hate speech online have already reached the courts. It is reasonable to assume, however, that the vast majority of hate speech incidents that have happened or are happening in the Internet never reaches courts or any other similar institution. So what are the ways to combat hate speech online, in particular, in major social media? Are they effective and how do they correspond to the case law of the European Court of Human Rights? Who decides what hate speech on the Internet is? These and other questions are discussed in this chapter through the analysis of the guidances and practise of the major social media – Facebook, Google (YouTube) and Twitter.

As was noticed in one of the recent studies on fake news – another topic of current interest, also closely connected to the freedom of expression – social media in western countries enjoy not being held responsible for the content that their users publish on the platforms (Niklewicz, K., 2017, p.29). According to the author of this study, this is a result of a mutual desire of the US and the EU to protect industry without full understanding of the potential of these new media (ibid). In other words, this legal vacuum creates a situation when social media themselves are able to decide what content they want to prohibit or restrict and how they want to do it.

When it comes to the online hate speech regulation, it was decided to analyse community standards, guidances, users' rules and even official statements as a primary source of information regarding the positions of the respective social media.

5.1. Hate speech definitions in social media's guidances

Hate speech in Facebook

The primary source for Facebook's hate speech policy is Facebook Community Standards that describe the policy and rules of this social media. Among others, in the Facebook Community Standards sections like "Violence and Criminal behaviour" and "Objectionable content" can be found. Interestingly enough, hate speech is assigned to the latter, whereas the former discusses issues such as violence, promoting crime, dangerous organizations etc. (see: Facebook Community Standards). Hate speech, on the other hand, is placed among "adult nudity/sexual activity" and "violence and graphic content" (ibid). This might be regarded as confusing, giving that hate speech is isolated by Facebook in a separate category from criminal activity and violence, even though it can be closely related issues.

In the Facebook Community Standards it is stated that the social media understands under hate speech a content that directly attacks people based on specific characteristics such as race, nationality, religion, sex, gender, sexual orientation etc. (Facebook Community

Standards, section Hate speech). It is evident that this definition closely corresponds to the definitions of hate speech discussed in the chapter 2.2 of the present thesis. Moreover, Facebook's list is even broader and includes additional categories.

In a separate section of the social media's help centre - Tools for addressing abuse – Facebook encourages its users to use special tools that were implemented to avoid abusive content. For example, Facebook advices to unfriend or block a person who sent an offensive message, use privacy settings and report the abuse and abuser (Facebook, Tools for addressing abuse). Report link is provided near the content on Facebook, which makes this procedure relatively easy to use.

Finally, in an additional statement, Facebook's VP of Global Public Policy Marne Levine claims that Facebook prohibits content that is harmful – such as anything that can result in real violence in the real world or anything that emotionally distresses an individual – but allows controversial or offensive content (Levine, 2013). Whereas this seems to be in accordance with the Article 10 and with the case law of the European Court discussed above, it is not clear how exactly Facebook's content moderators separate offensive content from the content that distresses its users. In reality, this boundary is so blurred that even the European Court quite often takes its decisions not unanimously, but with a number of dissenting opinions. How the social media can promise to distinguish one from another, remains an open question.

In fact, the responses to this statement clearly indicate that in practice the rules and policy of the social media do not function that smoothly as promised. Thus, the statement by Marne Levine was published in 2013 and subsequently has received more than 2,400 comments. Many users criticize Facebook for its policy and actions regarding hate speech and demonstrate alleged mistakes and passive approach towards the reports of abusive content (see commentaries in Levine, 2013). Whereas it is not possible to prove all the users' claims and published links, content analysis of the comments clearly indicates that Facebook's approach does not fully solve the problem of hate speech on this platform.

This is also evident from the two studies conducted by German Ministry of Justice in 2016 and 2017. In 2016, the study of Facebook's practice regarding hate speech regulation has shown that only 46% of flagged criminal content was deleted by the social media, out of which 42% – during the promised 24 hours (Löschung rechtswidriger Hassbeiträge bei Facebook, 2016). On the other hand, when the rest was later reported via direct email at the Facebook support team, 84% of criminal content was deleted, 48% - during the first 24 hours (ibid).

In 2017 the results of the same study were even more alarming: only 39% of flagged criminal content was deleted by Facebook, 33% – during the first 24 hours (Löschung rechtswidriger Hassbeiträge bei Facebook, 2017). After the remaining hate speech was reported via email, 88 percent of criminal content was deleted, 76 percent – within 24 hours (ibid).

That is why organizations such as the International Auschwitz Committee expressed their frustration and dissatisfaction with the practice of Facebook regarding hate speech. The Committee in particular stated that Facebook displays signs of arrogance towards this problem and thus contributes "to poisoning the social climate" especially in Germany, but also in the world in general (Deutsche Welle, 2016).

Hate speech in YouTube (Google)

One of the most popular platforms in the Internet – YouTube – also frequently encounters hate speech published by its users. YouTube, which belongs to the Google Company, has its own policy regarding this issue.

Thus, in the Policy center of the site a specific section "Hate speech policy" can be found. In this section YouTube defines hate speech as content that promotes violence or incite hatred against groups or individuals based on such characteristics as race, ethnicity, religion, disability, age, gender, sexual orientation or veteran status (YouTube, Policies, safety, and reporting). These categories can be compared to the ones described by Facebook: whereas there are some minor differences, such as mentioning of the veteran status by YouTube, in general the definitions may be considered similar.

YouTube, too, proposes its users several possibilities to combat hate speech on the platform: block the user, flag the video, file an abuse report in case numerous violation by the same user are found or report a violation of the local laws. Both flagging and

reporting tool are easily acceptable and convenient to use (YouTube, Report inappropriate content).

.

However, another study conducted by the German Ministry of Justice estimated that in 2016 YouTube deleted only 10% of the flagged content (Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2016). Direct report by a user with so-called privileged account (possible only on YouTube and Twitter, but not on Facebook) has led to the removing of another 35% of the criminal content (ibid). Finally, after direct report of abuse content via email, YouTube deleted 53% of such content, leaving only 2% online (ibid).

The results were better in a similar study in 2017. YouTube has significantly improved its response towards flagged content and deleted 90% percent of it (Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2017). After a direct email to the support center, the remaining 10% of criminal content were also removed from the platform.

Hate speech in Twitter

In its turn, Twitter has its own rules that describe content boundaries and limitations connected to the abusive behaviour. Twitter does not use the term “hate speech”, but mentions hateful conduct alongside with violent threats, harassment, multiple account abuse, self-harm impersonation, etc. (Twitter rules)

Twitter's rules define hateful conduct as promotion of violence against people based on race, ethnical or national origin, sexual orientation, gender, religion, age, disability or disease – with a slight difference, the same categories that were mentioned in the Facebook Community Standards (Twitter Rules). Thus, it is possible to conclude that although there is no universally accepted definition of hate speech in the normative acts, all three platforms that are analysed in the present thesis define hate speech or hateful conduct via similar definitions.

Nevertheless, the study commissioned by the German Ministry of Justice, quoted above, estimates that in 2016 Twitter deleted only 1% of flagged criminal content (Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2016). Another 75% were removed after direct report by user with a privilege account. Finally, 6% were deleted after contact via email to the support center, leaving 18% of reported criminal content online (ibid).

According to a similar study in 2017, Twitter did not improve its practice regarding flagged criminal content: only 1% was removed (Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2017). However, direct report and contact via email to the support center proved to be more effective: eventually, 100% of the criminal content were deleted by Twitter (ibid).

To sum it up, definitions of hate speech in three major social media are presented in Table 3:

Table 3. Hate speech definitions by Facebook, YouTube and Twitter

Facebook	Content that directly attacks people based on specific characteristics such as race, nationality, religion, sex, gender, sexual orientation, disability, etc. Attempts at "humor or satire" that might be perceived as "threat or attack" are allowed. (Facebook Community Standards, section Hate speech).
YouTube (Google)	Content that promotes violence or incites hatred against groups or individuals based on such characteristics as race, ethnicity, religion, disability, age, gender, sexual orientation or veteran status (YouTube, Policies, safety, and reporting).
Twitter	No mentioning of the term "hate speech", but hateful conduct that is defined as "promotion of violence" against people based on race, ethnical or national origin, sexual orientation, gender, religion, age, disability or disease (Twitter rules).

5.2. Regulation of hate speech on social media

Social media employ three main strategies in regulating hate speech on the platforms: use of automatic regulation[5], reports submitted by users and content moderation that is done by its content-reviewers (Laub, 2019). At the end of 2018 Facebook alone employed 15 000 content-reviewers who moderate content in more than 50 languages (Facebook Newsroom, 2018).

Already in 2015 three major technological companies – Facebook, Google and Twitter - stated that they were going to delete hate speech from their platforms within 24 hours (Reuters, 2015). This decision was taken amid the so-called refugee crisis in 2015, as an additional measure to combat rising online racism.

Just a year later, in 2016 the European Commission issued a Code of Conduct on countering illegal online hate speech (European Commission, 2016). It is a voluntary code that demands to remove the majority of hate speech within 24 hours, while defining hate speech as a conduct that is "publicly inciting to violence or hatred" towards groups or individual members of a group that is distinguished by special characteristics such as race, religious beliefs, nationality or ethnicity" (ibid).

Facebook, Twitter, YouTube, and Microsoft all signed the Code of Conduct in May 2016 (European Commission, 2017). In 2017, European Commission published a subsequent

[5] Artificial intelligence will be discussed below in paragraph 5.3.

press-release in which it underlined an important progress in combating hate speech online. However, challenges remain significant: according to the European Commission's data, in 2016-2017 IT companies

- removed hateful content only in 59% reported cases;
- reviewed 51% of the reported hateful content within 24 hours (ibid).

Finally, in the fourth review of the implementation of the Code of Conduct, published in 2019, European Commission concluded that hateful content flagged by users

- was deleted by IT companies in 72% of cases,
- was reviewed within 24 hours on average in 89% of cases (European Commission, 2019, p.1).

Altogether, during a period of 6 weeks in November-December 2018 the IT companies that signed the Code of Conduct[6] received 4392 notifications of the hateful content from the countries-members of the EU (ibid).

According to the Commission's review, deletion of the content by IT companies depends on its nature: content that advocates for violence against specific groups of people was removed in 85,5% of the cases, whereas the defamatory speech and images were deleted only in 58,5% of the cases (ibid, p.3). The most common types of hate speech were hate speech based on nationality (including speech against migrants), hate speech based on

[6] Apart from the above mentioned Facebook, Twitter, YouTube, and Microsoft, Instagram (279) and Google+ have joined the Code of Conduct in 2018 (European Commission, 2019, p.2).

sexual orientation and hate speech based on religious believes, specifically – anti-Muslim speech (ibid, p.5).

Independently from the European Commission and taking into account that percentage of removal of the flagging content was disproportionately weak in 2016-2017, in June 2017 German Parliament has voted to adopt a Social Networks Enforcement Law that is supposed to make social media more accountable for the content posted by their users. This law requires social media to pay up to 50 million euros in case they fail to remove hate speech from their platforms. Similar propositions can be heard in France and Great Britain (Toor, 2017).

Whereas this measure may be seen as a resolute and strong attempt to combat hate speech in the major social media, some organizations expressed their concerns regarding the effect of this law on the freedom of speech in Germany. For instance, David Kaye, UN Special Rapporteur to the High Commissioner for Human Rights, noticed that the German law gives private firms too much responsibility to police freedom of expression (McGoogan, 2017). According to Kaye, legislation, capable of limiting free speech, must be applied by an independent body (ibid).

Some other experts expressed their concerns regarding this measure. Mirko Hohmann, a Project Manager at the Global Public Policy Institute (GPPi) claimed that such a law will make social media companies delete content "excessively", after they facing fines of up to 50 million euros (McGoogan, 2017). This may lead to errors and removal of lawful

content as a precaution (ibid). Whereas the influence of the German law on hate speech and freedom of speech is still to be examined, there are other ways to combat hateful content online.

5.3. Alternative ways of combating online hate speech: attempts of automatic regulation

In recent years a new approach towards online hate speech regulation has been developed – namely, automatic regulation of such content. Artificial intelligence means that hate speech is being regarded from the point of view of computer science. In 2018-2019 this approach is still a relatively new field and available tools for automatic hate speech regulation are scarce, among other things – due to lack of systematic data (Fortuna, Nunes, 2018, p.3). But the number of articles on automatic hate speech regulation from computer science and engineering is rising in recent years, which indicates that technologies are being developed (ibid, p.22).

Algorithms for detecting hate speech are based on text mining and machine learning technologies that attempt to classify hate speech (ibid, p.16). Interestingly enough, researchers from the computer science emphasize that one of the problems in conducting this classification is low level of agreement regarding hate speech classification "by humans" (ibid, p.25).

While acknowledging that automatic regulation of hate speech remains an open issue, some researchers report positive results of binary classification "hate speech-not hate speech" (Del Vigna et al, 2017, p.94). Taking into account the huge amount of user-generated content, computer scientists underline that human moderators are not capable of monitoring major social media, and thus automatic regulation has a great potential for the future (ibid).

As was mentioned above, Facebook, YouTube and Twitter all employ automatic content regulation that to a certain extent concerns hate speech, but the availability of data regarding this regulation, its accuracy, objectivity and further consequences of such approach remain an issue for another study. In the current thesis it is important to notice that currently social media also rely on human moderators who apply the rules specified in the guidances inconsistently (Tobin et al, 2017). Facebook seems to acknowledge this problem, with Vice President of the company Justin Osofsky saying to journalists "we must do better" (ibid). At the same time, in 2019 Mark Zuckerberg called for a "global regulation" that could establish content and data standards (Laub, 2019).

5.4. Other initiatives to counteract online hate speech

Finally, one more approach related to online hate speech regulation must be mentioned – initiatives, independent from governments or social media. Most prominent examples of such initiatives is "Hate Speech International".

“Hate Speech International” is an independent network of journalists who conduct cooperative and international research on extremism and hate speech online and offline. The initiative was founded by a Norwegian journalists Kjetil Stormark and Øyvind Strømmen and received grants from Norwegian Freedom of Expression Foundation (Fritt Ord) and Ministry of Foreign Affairs.

“Hate Speech International”, however, operates independently and aims at displaying radicalization and extremist networks (Hate Speech International, About). For this purpose “Hate Speech International” regularly conducts related studies and publishes them on their website.

“Hate Speech International” is not the only initiative to combat hate speech that can be found online. For instance, similar aims – to monitor, to report, to educate about hate speech and policies and tools against it – had a project “No hate speech youth campaign” that was conducted by the Council of Europe in 2012-2017, as well as presently active Italian project eMore and pan European project MANDOLA, funded by Rights, Equality and Citizenship (REC) Programme of the European Commission. of the European Union.

6. Conclusion

In the present thesis the collision between freedom of expression and hate speech has been explored from two different perspectives: the one of the European Court of Human Rights, established in 1959, and the one of the three major social media platforms, developed during the last decade. In particular, case law of the Court related to hate speech has been analysed, as well as community standards and users' rules of Facebook, YouTube and Twitter.

Several conclusions have been made during analysis of the case law of the European Court of Human Rights. The most important of them were:

1) European Court of Human Rights does not provide a single definite definition to the term "hate speech", preferring to leave it open in order to approach each case on an individual basis. This attitude very much corresponds to the general lack of a single definition of the term "hate speech", as was observed during the review of scientific literature. Existent definitions, discussed in the paragraph 2.2., permitted to construct following broad and flexible hate speech definition for the purposes of the present thesis: we propose to understand under hate speech, utterances that (1) encourage hatred (2) towards individuals or groups of people (3) based on various particular characteristics (4) which are stigmatized and (5) are employed to legitimize hostility. Furthermore, hate speech (6) may be understood as a continuum that (7) does not necessarily result in violence, but (8) silences, discriminates and threatens targeted individuals or groups;

2) The term “hate speech” was not included in the European Convention on Human Rights. European Court of Human Rights first used this term in four judgments, passed in 1999, namely in Sürek v. Turkey (No. 1, par. 62); Sürek & Özdemir v. Turkey (par.63); Sürek v. Turkey (No. 4, par. 60) and Erdogdu & Ince v. Turkey (par. 54). Without providing an exact definition of the term, in the fourth of these cases the Court stressed that whereas some information may lawfully “offend, shock or disturb”, it the case Sürek v. Turke the utterance in question was “hate speech and glorification of violence” (Sürek v. Turke, 1999, par.62). In other words, the Court emphasized the fact that there is a new type of speech – expression that goes beyond the protection given to the offensive speech by the judgment in Handyside v. the United Kingdom, 1976, and specifically underlines violence as a necessary component of it. Yet, even after the introduction of the term “hate speech” in its practice, the Court quite often omits it and uses instead various synonymous phrases;

3) In 2019 the Court published an updated version of its Factsheet hate speech where it provides a typology of cases related to hate speech. This typology seems to be unnecessary detailed and not completely straightforward, yet it provides a full overview of the case law that concerns with information that the Court itself qualifies as hate speech. One could assume that such approach stresses an idea that it is not possible to provide a single finite definition to hate speech which instead should be regarded as an open spectrum of speech, even if some experts emphasize a need not to expand the term “hate speech” indefinitely (Tulkens, 2012, p.295);

4) Without a definition, case law of the Court remains the main source for scholars to analyse the Court's approach towards hate speech. The analysis of the case law, indeed, illustrates how the Court determines whether the speech in question is hate speech or not. Thus, the Court seems to employ several criteria on a case to case basis:

a) Nature of the speech;

b) Purpose of the speech (incitement of hatred, violence, and fear as opposite to starting public debate;

c) Significance of the speech (presence or absence of public interest);

d) Context of the speech;

e) Necessity of convincing the speech and the speaker in a democratic society;

f) Presence of a balance between freedom of expression and rights of others in decisions made by the national courts.

5) Additionally, following argumentation of the Court was found in the analysed judgments from the period 2012-2019:

a) There should be a balance between the protection of individual rights and public interest in freedom of expression (Aksu v. Turkey, 2012);

b) Necessity of restrictions of freedom of information must be established "convincingly" (ibid, par.64);

c) Hate speech is connected by the Court to inciting violence by "appealing to the irrational hatred against a person or a group of people" (Öner and Türk v. Turkey, 2015, par.24);

d) Critical assessment of the authorities without incitement to violence does not constitute hate speech (ibid);

e) Racist verbal abuse is highly relevant in cases involving ethnic minorities and states have an obligation to establish possible racist motive behinds attacks (Balázs v. Hungary 2015, par.53);

f) Public statements that concern a matter of public interest must have higher protection under the Article 10 (Perinçek v. Switzerland, 2015, par. 197);

g) On the contrary, speech with commercial purposes – such as commercial advertisement – should have less protection by the Article 10 than political debate (Raelien Suisse v Switzerland, 2012, par.62);

h) Religious fundamentalism coupled with incitement to violence amounts to hate speech, as much as attack on people based on their religious beliefs (Belkacem v. Belgium, 2017, par.34);

i) Even if certain statements do not propagate hatred but are allegations that reinforce prejudices, they might be prohibited (Vejdeland and Others v. Sweden, 2012, 54);

k) Discrimination based on sexual orientation should be considered as serious as discrimination on the ground of race, sex, colour or origin (Vejdeland and Others v. Sweden, 2012, par.42). Homophobic motives must be investigated (Identoba and Others v. Georgia, 2015).

l) Internet news portals have "duties and responsibilities" when their users disseminate unlawful speech through their sites (Delfi AS v. Estonia, 2015, par.113);

m) Even hypothetical statements might be prohibited, taking into account the context (Smajić v. Bosnia and Herzegovina, 2018, par.39);

n) Criminal conviction is disproportionate for offensive speech without incitement to hatred (Savva Terentyev v. Russia, 2018, par.86).

At the same time, subsequent analysis of the social media's approach towards hate speech provided the following observations:

6) Taking into account an amount of user-generated content, it is possible to assume that the majority of hate speech incidents now happens online and never reaches courts. Three major social media – Facebook, YouTube (Google), Twitter – all have developed community standards and users' rules where they define hate speech and prohibit it on its platforms. Whereas there are some minor differences between the definitions of hate speech by these companies, it is possible to acknowledge that they are more or less identical. The problem lies in the implementation of these rules;

7) For instance, Facebook claims to distinguish "offensive content" from the content that "distresses its users". This is a daunting task even for the European Court that quite often takes its decisions not unanimously, but with a number of dissenting opinions. How exactly and how accurately social media approach this issue, remains an open question. More transparency from social media is needed in order to investigate this issue;

8) Several studies conducted by German Ministry of Justice and by the European Commission in recent years indicated that social media failed to remove 100% of content flagged as hate speech, although this statistic improves year after year;

9) In western countries, social media are not being held responsible for the content published by their users. That means that social media themselves decide what contact they want to prohibit or restrict. Yet, there is a tendency in the opposite direction. In 2016 the European Commission issued a Code of Conduct on countering illegal online hate speech that was signed by Facebook, YouTube (Google), Twitter, among others. Simultaneously, in June 2017 German Parliament has voted to adopt a Social Networks Enforcement Law that declares social media responsible for illegal hateful content on their platforms;

10) This law has been criticised by experts who believe that it gives too much responsibility to police freedom of expression to the private companies. Among others, this concern was expressed by UN Special Rapporteur to the High Commissioner for Human Rights David Kaye who urged instead to organise an independent body for online content regulation. Interestingly enough, Facebook CEO Mark Zuckerberg recently called for a "global regulation" of content and data, too (Laub, 2019);

11) Eventually, at the present moment online hate speech regulation lies with social media that employ three strategies: automatic regulation, users' reports and content moderation done by human reviewers. In 2018 Facebook alone employed 15 000 content-reviewers. These people, de-facto, serve as new judges in a new court. Their qualification for this task and consistency with established practice, such as, for instance, the practice of the European Court of Human Rights, are highly questionable;

12) Meanwhile, a new approach towards online hate speech regulation is being developed by computer scientists who believe that human moderators cannot monitor major social media due to the amount of user-generated content. Automatic content regulation is based on text mining and machine learning technologies that attempt to classify hate speech. Whereas automatic regulation might result in decreasing of online hate speech, it is reasonable to assume that application of this technology will create new threats and new restrictions for freedom of expression. In reality, it means that society is going to substitute already not sufficiently qualified human moderators with self-educated automatic technologies that will decide for us what content is allowed for humans and what it judges necessary to prohibit;

13) Overall, this recent popularity of extensive regulation approach seems to indicate that human civilisation failed to take its hatred under control. Instead of promoting alternative measures – such as improving education, rising awareness about the problem, increasing tolerance and intercultural dialogue, providing voice to the victims of hate speech – society came to a conclusion that prohibition and automatic recognition will conclusively solve the problem, as if removal of online hate speech will at the same time remove hatred from the offline communities.

Automatic hate speech recognition is, indeed, needed for a vast amount of user-generated content; but without supervision by international qualified body, it cannot guarantee freedom of expression, as much as without broad educational and cultural measures,

restrictions imposed by social media alone cannot possibly protect human rights in the long-term perspective.

7. References

Articles and books

Abrams. (2012). On American Hate Speech Law / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P. Cambridge Univ. Press, p.116-129.

Baker, C. E. (2012) Hate speech / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P. Cambridge Univ. Press, p.57-80.

Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, Ill: The Free Press.

Bleich , E. (2014). Freedom of Expression versus Racist Hate Speech: Explaining Differences Between High Court Regulations in the USA and Europe. *Journal of Ethnic and Migration Studies*, 40:2, p. 283-300.

Bonotti, M. (2017). Religion, hate speech and non-domination. *Special Issue: Religion and Public Life*, Vol. 17(2), p. 259–274.

Buyse, A. (2014). Dangerous expressions: The ECtHR, violence and free speech. *ICLQ vol 63*, pp 491–503.

Davids, N. (2018). On the (in)tolerance of hate speech: does it have legitimacy in a democracy?, *Ethics and Education*, 13:3, p. 296-308.

Del Vigna, F., Cimino, A., Dell'Orletta, F., Petrocchi, M., Tesconi, M. (2017). Hate me, hate me not: Hate speech detection on Facebook. *Proceedings of the 1st Italian Conference on Cybersecurity*, p. 86–95.

Downs, D. M., Cowan, G. (2012). Predicting the Importance of Freedom of Speech and the Perceived Harm of Hate Speech. *Journal of Applied Social Psychology*, 42, 6, p.. 1353–1375.

D'Souza, T., Griffin, L., Shackleton, N., Walt, D. (2018). Harming Women with Words: The Failure of Australian Law to Prohibit Gendered Hate Speech, *41 U.N.S.W.Law Journal*. 939, p.939-976.

Fortuna, P. Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv. 51, 4, Article 85*.

Gelber, K, McNamara, L. (2016). Evidencing the harms of hate speech/ *Social Identities*, 22:3, p. 324-341.

Greene, J. (2012) Hate speech and the Demos / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P., Cambridge Univ. Press, p.92-116.

Hall, M. A., Wright R. F. (2008). Systematic Content Analysis of Judicial Opinions. *California Law Review*, Vol. 96, No. 1, p. 63-122.

Haraszti, M. (2012). Foreword: hate speech and the Coming Death of the International Standard before it was born / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P., Cambridge Univ. Press ; p. xiii-xix.

Holoubek, M., Kassai, K., Traimer, M. (2014). *Grundzüge des Rechts der Massenmedien*, Verlag Österreich.

Holsti, O.R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.

Korn, G. (2014). *Einfuhrung in das Kommunikationsrecht*. Facultas Verlags- und Buchhandels AG, p. 49-50.

Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage. (2d ed., 2004).

Maitra, I, McGowan, M. K. (2010). On Racist Hate Speech and the Scope of a Free Speech Principle. *The Canadian Journal of Law and Jurisprudence*, 23, p 343-372.

Mayring, P. (2000) Qualitative Content Analysis. *Forum Qualitative Sozialforschung, Volume 1, No. 2, Art. 20*.

McGonagle, T. (2013). The Council of Europe against online hate speech: Conundrums and challenges. *MCM; No. 2013(005)*. Belgrade: Republic of Serbia, Ministry of Culture and Information.

Molnar, P. (2012). Responding to "Hate Speech" with Art, Education, and the Imminent Danger Test"// In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P. Cambridge Univ. Press, p.183-198.

Niklewicz, K. (2017). *Weeding out Fake News: An Approach to Social Media Regulation*. Brussels, Wilfried Martens Centre for European Studies. Retrieved from: https://martenscentre.eu/publications/weeding-out-fake-news-approach-social-media-regulation

Parekh, B. (2012). Is there a case for banning hate speech? / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P., Cambridge Univ. Press ; p.37-57.

Richardson-Self, L. (2017). Woman-Hating: On Misogyny, Sexism, and Hate Speech. *Hypatia vol. 33, no. 2,* p. 256-272.

Richter A. (2012). One step beyond hate speech / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P., Cambridge Univ. Press, p.290-306.

Schauer, F. (2012). Social Epistemology, Holocause Denial, and the Post-Millian Calculus / In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P., Cambridge Univ. Press ; p.129-144.

Sottiaux, S. (2011). 'Bad Tendencies' in the ECtHR's 'Hate Speech' Jurisprudence. *European Constitutional Law Review,* 7, p. 40–63.

Sparker A. (2005). Narrative analysis: exploring the whats and hows of personal stories. *Holloway I (ed.).Qualitative Research in Health Care (1st edn).* Berkshire: Open University Press, p. 191–208.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation,* 7(17).

Suk, J. C. (2012). Denying Experience: Holocaust Denial and the Free-Speech Theory of the State/ In: *The content and context of hate speech. Rethinking Regulation and Responses* |ed. by| Herz, M., Molnar, P. Cambridge Univ. Press, p.144-164.

Tulkens, F. (2012). When to say is to do: Freedom of expression and hate speech in the case-law of the European Court of Human Rights. In: *Freedom of Expression: Essays in honour of Nicolas Bratza /ed.by/* Casadevall, J., Myjer, E., O'Boyle M., Austin, A. Oisterwijk, The Netherlands, Wolf Legal Publishers, p. 279-295.

Yong, C. (2011). Does Freedom of Speech Include Hate Speech? *Springer Science+Business Media B.V., Res Publica, 17,* p. 385–403.

Vaismoradi, M., Turunen, H. (2013). Content analysis and thematic analysis: Implications forconducting a qualitative descriptive study. *Nursing and Health Sciences*, ,15, p. 398–405.

Weber, R. P. (1990). *Basic Content Analysis*, 2nd ed. Newbury Park, CA.

Reports, declarations and other normative acts

Council of Europe. (1997). *Recommendation No. R 97(20) of the Committee of Ministers of the Council of Europe to Member States on "hate speech"*. Retrieved from:

http://www.coe.int/t/dghl/standardsetting/hrpolicy/other_committees/dh-lgbt_docs/CM_Rec(97)20_en.pdf

Council of Europe. (2013). *Report «The hate factor in political speech: Where do responsibilities lie?»*. Retrieved from: https://rm.coe.int/16800c170e

European Convention on Human Rights. (1950). Retrieved from: https://www.echr.coe.int/Documents/Convention_ENG.pdf%23page=9

European Commission. (2016). *Code of Conduct on countering illegal online hate speech*. Retrieved from: http://europa.eu/rapid/press-release_IP-16-1937_en.htm

European Commission. (2017). *Countering online hate speech. Commission initiative with social media platforms and civil society shows progress*. Retrieved from: http://europa.eu/rapid/press-release_IP-17-1471_en.htm

European Commission. (2019). *Factsheet - 4th monitoring round of the Code of Conduct*. Retrieved from: https://ec.europa.eu/info/files/factsheet-4th-monitoring-round-code-conduct_en

European Court of Human Rights (2019). Factsheet hate speech. Retrieved from: http://www.echr.coe.int/Documents/FS_hate_speech_ENG.pdf

International Convention on the Elimination of All Forms of Racial Discrimination. (1965). Retrieved from: http://www.ohchr.org/EN/ProfessionalInterest/Pages/CERD.aspx

International Covenant on Civil and Political Rights. (1966). Retrieved from: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

OSCE. (2005). *Combating hate crimes in the OSCE region: an overview of Statistics, Legislation, and National Initiatives*. Organization for Security and Co-operation in Europe. The Representative on Freedom of the Media.

OSCE (2009). *Hate Crime Laws. A practical Guide*. Organization for Security and Co-operation in Europe, The Representative on Freedom of the Media.

OSCE (2009). *Preventing and responding to hate crimes. A resource guide for NGOs in the OSCE region*. Organization for Security and Co-operation in Europe, The Representative on Freedom of the Media.

Universal Declaration of Human Rights. (1948). Retrieved from: http://www.un.org/en/documents/udhr/

UN Committee on the Elimination of Racial Discrimination (CERD). (2013). *General recommendation No. 35 : Combating racist hate speech*, CERD/C/GC/35. Retrieved from: https://www.refworld.org/docid/53f457db4.html [accessed 13 May 2019]

Case law of the European Court of Human Rights

(Access via hudoc.echr.coe.int):

Aksu v. Turkey, 2012

Balázs v. Hungary 2015

Balsytė-Lideikienė v. Lithuania, 2008

Belkacem v. Belgium, 2017

Delfi AS v. Estonia, 2015

Dink v. Turkey, 2010

Erbakan v. Turkey, 2006

Erdogdu & Ince v. Turkey, 1999

Fáber v. Hungary, 2012

Féret v. Belgium, 2009

Garaudy v. France, 2003

Glimmerveen and Haqenbeek v. the Netherlands, 1979

Gündüz v. Turkey, 2003

Handyside v. United Kingdom, 1976

Hösl-Daum and Others v. Poland, 2014

Ibragim Ibragimov and Others v. Russia, 2018

Identoba and Others v. Georgia, 2015

Ivanov v. Russia, 2007

Jersild v. Denmark, 1994

Király and Dömötör v. Hungary, 2017

Lehideux and Isorni v. France, 1998

Le Pen v. France, 2010

Leroy v. France, 2008

Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, 2016

M'Bala M'Bala v. France, 2015

M.C. and A.C. v. Romania, 2016

Nix v. Germany, 2018

Norwood v. United Kingdom, 2004

Öner and Türk v. Turkey, 2015

Pavel Ivanov v. Russia, 2007

Perinçek v. Switzerland, 2015

Pihl v. Sweden, 2017

Raelien Suisse v Switzerland, 2012

Roj TV A/S v. Denmark, 2018

Schimanek v. Austria, 2000

Šimunić v. Croatia, 2019

Smajić v. Bosnia and Herzegovina, 2018

Soulas and Others v. France, 2008

Stern Taulats and Roura Capellera v. Spain, 2018

Stomakhin v. Russia, 2018

Sürek v. Turkey (No. 1), 1999

Sürek & Özdemir v. Turkey, 1999

Sürek v. Turkey (No. 4), 1999

Vejdeland & others v. Sweden, 2012

Web-Resources

European Union Agency for fundamental rights, Hate Speech and Hate Crimes Against LGBT Persons. Access via:

http://fra.europa.eu/sites/default/files/fra_uploads/1226-Factsheet-homophobia-hate-speech-crime_EN.pdf

Facebook Community Standards. Access via:

https://www.facebook.com/communitystandards

Facebook Newsroom. (2018). Facts About Content Review on Facebook**.** Access via:

https://newsroom.fb.com/news/2018/12/content-review-facts/

Facebook. Tools for Addressing Abuse. Access via:

https://www.facebook.com/help/tools

Hate Speech International. Access via:

https://www.hate-speech.org/about/political-backdrop/

Hate speech Monitor. Access via:

https://www.hate-speech.org/hatespeech-monitor/

Löschung rechtswidriger Hassbeiträge bei Facebook, 2016. Access via:

http://www.fair-im-netz.de/WebS/NHS/SharedDocs/Downloads/DE/09262016_Testergebnisse_Facebook.pdf?__blob=publicationFile&v=2

Löschung rechtswidriger Hassbeiträge bei Facebook, 2017. Access via:

http://www.fair-im-netz.de/WebS/NHS/SharedDocs/Downloads/DE/03142017_Recherchebericht_Facebook.pdf?__blob=publicationFile&v=3

Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2016. Access via:

https://www.bmjv.de/SharedDocs/Downloads/DE/Artikel/09262016_Testergebnisse_jugendschutz_net_Hasskriminalitaet.pdf?__blob=publicationFile&v=1

Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter, 2017. Access via:

http://www.bmjv.de/SharedDocs/Downloads/DE/Artikel/03142017_Monitoring_jugendschutz.net.pdf?__blob=publicationFile&v=1

MANDOLA project. Access via:

http://mandola-project.eu/

No hate speech youth campaign. Access via:

https://www.coe.int/en/web/no-hate-campaign

Project eMore. Access via:

https://www.emoreproject.eu/

Statista. Most popular social networks worldwide as of April 2019. Access via:

https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Twitter Rules. Access via:

https://support.twitter.com/articles/18311

YouTube, Policies, safety, and reporting. Access via:

https://support.google.com/youtube/answer/2801939?hl=en

YouTube, Report inappropriate content. Access via:

https://support.google.com/youtube/answer/2802027

Press articles

Deutsche Welle. (2016). International Auschwitz Committee accuses Facebook of arrogant approach to hate speech. Retrieved from http://www.dw.com/en/international-auschwitz-committee-accuses-facebook-of-arrogant-approach-to-hate-speech/a-36787344

Laub, Z. (2019). Hate Speech on Social Media: Global Comparisons.Council on foregn relations. Retrieved from: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

Levine, M. (2013). Controversial, Harmful and Hateful Speech on Facebook. Retrieved from: https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054

McGoogan, C. (2017). Germany to fine Facebook and YouTube €50m if they fail to delete hate speech. Retrieved from: http://www.telegraph.co.uk/technology/2017/06/30/germany-fine-facebook-youtube-50m-fail-delete-hate-speech/

Reuters. (2015). Facebook, Google, Twitter agree to delete hate speech in 24 hours. Retrieved from: http://www.reuters.com/article/us-germany-internet-idUSKBN0TY27R20151215

Tobin, A,, Varner, M, Angwin, J. (2017). Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up. Pro Publica. Retrieved from: https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes

Toor, A. (2017). Germany wants to fine Facebook over hate speech, raising fears of censorship. Retrieved from: https://www.theverge.com/2017/6/23/15852048/germany-hate-speech-facebook-twitter-fine-censorship

8. Abstract

One of the most essential problems of modern media law is the collision between freedom of expression and hate speech. Rapid development of the Internet and social media have greatly contributed to the realization of freedom of expression worldwide, but at the same time increased dissemination of hate speech. In the new media environment, establishing control over hate speech while simultaneously protecting freedom of expression became a global issue.

This thesis attempts at looking at this problem from two different perspectives: from the position of the European Court of Human Rights which has a long history of passing judgments in cases related to hate speech, and from the point of view of the major social media which in the last decade have developed their own rules and measures regarding this issue.

In order to establish how the European Court of Human Rights decides whether the speech in question is hate speech or not, more than 30 judgments of the Court are analyzed in detail. The most significant arguments from the judgments are summarised in two tables.

The Court's approach is then compared with the one of the social media that attempt to combat hate speech with the help of human moderators and automatic filtering systems. A number of observations resulted from this comparison are provided in the conclusion.

Keywords: freedom of expression, hate speech, European Court of Human Rights, case law, social media

9. Zusammenfassung

Eines der größten Probleme des modernen Medienrechts ist die Kollision von Meinungsfreiheit und Hassrede. Die rasche Entwicklung des Internets und der sozialen Medien hat weltweit zur Verwirklichung der Meinungsfreiheit beigetragen, gleichzeitig hat sie aber die Verbreitung von Hassreden verstärkt. In der neuen Medienlandschaft wurde die Kontrolle über Hassreden bei gleichzeitigem Schutz der Meinungsfreiheit zu einem globalen Thema.

Diese Masterarbeit versucht dieses Problem aus zwei verschiedenen Perspektiven zu betrachten: aus der Position des Europäischen Gerichtshofs für Menschenrechte, der seit langem die Fälle im Zusammenhang mit Hassrede beurteilt, und aus der Sicht der wichtigsten sozialen Medien, die im letzten Jahrzehnt ihre eigenen Regeln und Maßnahmen zu diesem Thema entwickelt haben.

Um festzustellen, wie der Europäische Gerichtshof für Menschenrechte entscheidet, ob es sich um eine Hassrede handelt oder nicht, werden mehr als 30 Urteile des Gerichtshofs im Detail analysiert. Die wichtigsten Argumente aus den Urteilen sind in zwei Tabellen zusammengefasst.

Der Ansatz des Gerichtshofs wird dann mit dem der sozialen Medien, die Hassreden mit Hilfe von Moderatoren und automatischen Filtersystemen zu bekämpfen versuchen, verglichen. Die Beobachtungen, die sich aus diesem Vergleich ergeben, werden in der Schlussfolgerung präsentiert.

Suchbegriffe: Meinungsfreiheit, Hassrede, Europäischer Gerichtshof für Menschenrechte, Rechtsprechung, soziale Medien