Hate Speech: The issue of Offensive Language

Yulia Razmetaeva, Olena Dashkovska, Marina Khaustova

Department of Theory and Philosophy of Law, Yaroslav Mudryi National Law University.

ABSTRACT

This study aims to use hermeneutic and anthropological philosophical and linguistic approaches to understanding hate speech and its online constituent, as well as to protect human dignity as a starting point for regulating hate speech. We have also used logical, formal legal and comparative legal methods to construct approaches to global threats to human liberties and balancing freedom of expression, comparing legal regulation and legal practice concerning hate language in various legal systems. Results showed that Hate speech proliferation in the online mode not only negatively affects groups or individuals, but also contributes to extinction of respectful and democratic discourse per se.

Keywords: hate speech, online hate speech, freedom of expression, legal regulation, legal systems, digital age

Introduction

Freedom of expression is a fundamental right, a value common to humanity and one of the foundational tenets of democracy. Attempts to curtail this freedom leads to a slowdown in social development, the stopping of a public dialogue on important issues and, ultimately, can lead us to authoritarianism and totalitarianism. At the same time, human dignity and non-discrimination require equal and respectful treatment of people and groups, regardless of any distinctive characteristics, and the expression of opinions must not break human rights and contribute to the growth of hateful ideologies.

The nature, regulation and restrictions on hate language are considered across these issues, which is especially important in the digital age, when the dissemination of expressed opinions is akin to a wilderness fire—a sudden, running and difficult to stop.

In this investigation, we have utilized hermeneutic and anthropological philosophical ways to deal with understanding disdain discourse and its online constituent, just as to ensuring human nobility as a beginning stage for managing disdain discourse. We have additionally utilized consistent, formal lawful and relative lawful strategies to build ways to deal with worldwide dangers to human freedoms, adjusting opportunity of articulation and ensuring human respect, looking at lawful guideline and legitimate work on concerning disdain language in different overall sets of laws.

Hate Speech and the Problems of its Definition

Hate speech refers to forms of expression that are not or are not always protected by a freedom-of-expression. Firstly, the reason for this is the negative impact of hate speech. Thus, Alexander Tsesis points out that such a language seeks to puncture faith in democracy, since it encourages supporters to intolerant and inhumanly treat groups of people as unworthy of human rights and human dignity by nature (TSESIS, 2001).
The danger of the use of hate language can be expressed in the incultation of intolerance and discrimination, which includes the creation of a culture of hatred in society. The shattering effect of hate speech is amplified in the digital age by an extremely rapid and widespread dissemination of information. The adverse implications of terms of hatred can have an especially acute impact on conflict societies, which is further amplified in the digital age due to the information component of hybrid conflicts and global threats.

There is no generally accepted definition of language of hatred, although both definitions and criteria for recognizing hate speech can be found in international and national legal acts. Therefore, Article 19 of the International Covenant on Civil and Political Rights states that everyone has the right to adhere to his/her opinion without restriction, has the right to freedom of expression, although it is limited to “respect for the rights and reputation of others” or “the protection of state security, public order, health or morality of the population.” However, in accordance with Article 20 of the Covenant, any speech in favor of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence must be prohibited by law (International Covenant on Civil and Political Rights, 1966). Under Part 1 of Article 10 of the Convention relative to the Protection of Human Rights and Fundamental Freedoms, it is established that everyone has the right to freedom of expression. This right includes the freedom to hold opinions, receive and impart information and ideas, without government intervention and regardless of borders. At the same time, Part 2 of Article 10 of the Convention provides that the exercise of these freedoms, insofar as it involves duties and responsibilities, may be subject to certain formalities, conditions, restrictions or sanctions that are provided for by law and are necessary in a democratic society in the interests of national security, territorial integrity or public security for preventing disorder or crime, protecting health and morals, protecting the reputation or rights of others, preventing the disclosure of confidential information, or maintaining the authority and impartiality of the judiciary (Council of Europe, 1950). All of this attest to the fact that certain – hateful – forms of expression may not be protected by fundamental rights. First, this refers to such a significant and broad right as freedom of expression.

Council of Europe Committee of Ministers Recommendation № 97 (20) on hate speech provides that the latter includes all forms of expression that spread, incite, support or justify racial hatred, xenophobia, anti-Semitism and other forms of hatred caused by intolerance (Recommendation No. R (97) 20 of the Committee of Ministers to member states on “Hate Speech”, 1997). Consequently, hate speech embraces not only the language itself, but also images, videos, gestures, actions. In addition, hate speech can be called hate speech in relation to groups of people and to individuals.

As Anne Weber rightly pointed out, hate speech includes, first of all, rousing hatred against individuals or groups of individuals on the basis of their race, religious reasons, on the basis of differences between believers and non-believers, as well as incitement to other forms of hatred caused by intolerance, which manifest themselves in the form of aggressive nationalism and ethnocentrism (WEBER, 2010). These four characteristics – religion, race, nationality, and ethnicity – do not conclude the entire spectrum of forms of hatred, but represent the motives that tend to be encountered.

The two problems should be outlined when defining hate speech: (1) difficulty of qualifying certain forms of expression as hateful (further complicated by the lack of definition and situational criteria); (2) difficulty of qualifying the relevant forms of expression as not protected by the right to freedom of expression. An auxiliary feature can be the fact that hate speech is of no worth and does not carry any meaning, except for the actual expression of hatred.

Hate speech is part of language per se, but it has a negative connotation, including a strong and negative emotional connotation. As Vira Savondhak writes, “the very language cannot be adverse or well-meaning; it is only a subjective evaluation.” (SAVONCHAK, 2011). Such an evaluation is expressed with the help of emotions of rejection, anger, disgust, hatred. According to Elina Larinova, “hatred is an emotion with a pronounced negative-evaluative connotation (loathing, dislike, anger) when faced with a different system of values, identity, and culture. The object of hatred is a Stranger, and his/her strangeness is socially conditioned by a set of certain characteristics” (LARINA, 2010). The topic of a potential threat from a conditional other is often exploited in hate speech. As noted, “the carriers of some social identities, in order to achieve group unity, create and exploit social ideas about the negative image of the collective Other, that is a direct threat” (PUSHKAR & MALYSHEVA, 2015).

Thus, discriminatory and intolerant attitudes that are expressed through hate speech usually have two characteristics, namely: (1) use of distinctive features and (2) identification of an individual or group as “Other / Stranger” (RAZMETAeva, 2016). Thanks to the second characteristic, adherents of hate speech receive, at least in their minds, a moral right to an intolerant attitude. Subsequently, it is in danger of turning into such a delineation of identities, which provides for the labeling of a person or group as non-people as such creatures that are excluded from being humanely treated and, accordingly, any actions are permitted towards them.
Hate Speech and Its Online Component in the Digital Age

In the digital age the agenda of tolerance and non-discrimination, as well as freedom of expression, is becoming world spanning. On the one hand, online tools are capable of promoting diversity and openness, creating new ways of communication and forms of expression, allowing a wider variety of people to express their opinions in public. On the other hand, it becomes much more difficult to regulate and monitor hate speech than in the case of speech at rallies or in print media. All effective digital tools can work both for the betterment and for the detriment of universal human values.

As James Banks noticed, the exponential growth of the Internet, its anonymity, immediacy and global nature, is driving the spread of hate speech, including the growth of online hate groups and hate-related activities in cyberspace (BANKS, 2010). The Internet contains a huge number of websites that collect materials of a discriminatory and sharply hateful extremist nature. The possibility of being connected to the network, both in a very real sense and in the sense of almost continuous being in the information field, at least a superficial knowledge of the main events and news, have changed the ways of communication adopted between people and communities. In addition, social media are responsible for this in many respects. At the same time, the nature of these media, the growing individualization of content, customization based on personal preferences, algorithms for delivering results – all this has the effect of being inside a comfortable environment of a glass dome and with the feeling that everything you need is here, and outside there is an alien desert, possibly hostile. Social media can also contribute to simplifying the perception of information and polarizing society, therefore hate speech finds fertile ground here.

In a review on hate speech and terrorism, Naganna Chetty and Sreejith Alathur emphasize that both phenomena are closely related and makes use of online social networks (CHETTY & ALATHUR, 2018). A study of the main goals of hate speech in social media has shown that (1) the three main categories of online hate are the same – race, behavior and physical fitness, (2) the categories of behavior and physical aspects are more related to soft hate targets, such as fat or stupid people, which perhaps means targeting hate speech at groups of people that are usually not considered when documenting hate crimes (SILVA et al., 2016), that is, when documenting offline. In addition, the analysis of the flow and dynamics of messages on a huge dataset has confirmed that hateful content spreads further, wider and faster and has a greater reach than the users who do not have hatred (MATHEW et al., 2019). Addressing the specifics of online hate speech, Alexander Brown points out that online hate speech has already outperformed offline hate speech, measured as the share of total hate speech, and will continue to grow in the future (BROWN, 2018).

On the other hand, automated methods can find and classify the use of hate speech. This enables, for example, its escalation and extensive use to be assessed numerically (ZANNEITTOU et al., 2020). Lei Gao and Ruihong Huang propose the models of detection of hate speech using context (GAO & HUANG, 2017), which can be of particular importance in the digital age to differ hateful speech from sharp dispute, but factoring into public discourse on issues of significance.

In any case, in order to combat hate speech – online and offline – it is necessary to determine the boundaries beyond which the restriction of freedom of expression and other fundamental rights is permissible, the boundaries beyond which the violation of international and national legal acts begins. However, it is very difficult to draw a distinct line between them. Thus, in the practice of the European Court of Human Rights, one can find criteria for distinguishing between forms of expression that are not tolerant, but protected by the right to freedom of expression, and those that are not protected. In the well-known case Erbakan v. Turkey the European Court of Human Rights noted that tolerance and respect for the equal dignity of all people was the basis of a democratic pluralistic society, therefore, in some democratic societies it may be considered necessary to impose sanctions or prohibit any form of expression that disseminates, incites, encourages or justifies hatred based on intolerance, provided it is consistent with a legitimate aim (ERBAKAN, 2006). However, several rights that are equally protected by the Convention for the Protection of Human Rights and Fundamental Freedoms can compete with each other. The right to freedom of expression can be abridged with the right to freedom of thought, conscience and religion, the right to privacy, the prohibition of all forms of discrimination.

At the same time, if legal regulation includes the prohibition of hate speech, this could potentially jeopardize freedom of speech and lead to cessation of the expression of any unpopular ideas. Such regulation, in addition, potentially carries the risk of abuse and criminalization of certain forms of expression, and facilitates attempts to manipulate laws in order to stop discussion. Restrictions on expression to eliminate hate speech are also fraught with temptation to define what is inappropriate for public comment. Moreover, it can turn into a tool for suppressing dissatisfied voices. In addition, problems arise with the regulation of the online component of hate speech, since both the nature of the digital environment and the jurisdictional issues of cyberspace create additional obstacles in applying legal instruments here.
Problems of Legal Regulation of Hate Speech and Global Threats

The regulation of hate speech can become a dangerous tool in the hands of a censor, who is usually embodied in a public authority having a desire to prohibit some forms of expression. It is difficult, in turn, to dose censorship. In addition, paternalistic efforts to protect people from the negative effects of hate speech can reduce critical thinking and reduce open discussion. Should individuals and groups be given many opportunities for defending any beliefs consciously?

One of the points of view asserting that in the digital age such a field for freedom of expression as the Internet should not be regulated is built on the idea of its being an extraspacial space in which no state has or can have sovereignty (TSESSI, 2001). In addition, a significant proportion of the regulation of the digital space is no longer carried out by governments, but by corporations that own digital tools.

It is noted that since the regulation of hate speech is a particularly challenging task, it is possible that Internet companies should play a larger role in solving this problem on an everyday basis compared to government organizations and agencies (BROWN, 2018). However, this approach increases the risk of abuse by companies, because they are often driven by the pursuit of profit rather than considerations of fairness, and because companies may not have insight into legal values and principles. At the same time, business can contribute to solving the problem of hate speech in the online mode. Companies’ strong points could be issues related to multiple jurisdictions as well as technological realities such as mirror sites (ALKIVIADOU, 2019). At the same time, account must be taken of the differences between legal systems in area of the doctrine of regulation of freedom of expression and understanding of hate speech, as well as the inconsistency of legislation and the difference in judicial practice on these issues.

However, even in systems in which the uniformity of standards is ensured by unquestioningly recognized legal acts and judicial practice, there is a difference in the understanding of permissible and proportional measures regarding hateful forms of expression. Thus, in 2015, the ECHR in the case of Delphi AS v. Estonia concluded that bringing a commercial organization – the owner of a news portal – to moderate responsibility for the lack of an immediate response to the publication of comments under the article that clearly contain hate speech and incitement to violence, was not a disproportionate restraint of freedom of expression (DELFII AS, 2015). This was the first time that the ECHR resolved a complaint about liability for comments made by users on the site. However, the court decision contains a large number of reservations, in particular, the court noted the clearly illegal nature of the comments (obvious incitement to hatred and incitement to violence), the fact of the commercial basis and professional management of the news portal, the insufficient measures taken by the applicant company to immediately eliminate the comments and ensure real prospects for the authors of such comments to be held accountable, as well as the moderation of the sanctions imposed on the company. The judicial decision is regulated and applicable and does not apply to comments on social media. At the same time, the court noted that the automatic system for blocking comments due to the use of hate language in them in the form of offensive words was not enough.

It is unlikely that it will be possible to apply the requirements similar to traditional media for online hate speech from the point of view neither of expediency, nor from the point of view of regulatory capabilities. Moreover, the temptation to regulate it via legal prohibitions and censorship is faced with flexible, digital means of counteraction. At the same time, protecting human dignity and human rights online at the same level as offline requires the cooperation of all stakeholders to combat hate speech in the digital age. Such requirements are addressed, first, to states, as subjects of direct obligations in the field of human rights and those who can most effectively respond to global threats. In the digital age, such requirements may be imposed on corporations and international organizations, as the subjects having today influential digital tools, economic and political power. It is possible that in the future similar requirements will be addressed directly to the very individuals, and not through a general obligation not to violate the rights of others, given how much impact a flexible and dynamic digital space can give to individual users.

Conclusion

The rapid dissemination of information, including information containing hate speech, as well as the presence of global threats, makes one think about effective strategies for preventing and combating hate speech. Such strategies may comprise legal regulation and judicial practice, educational and information components.

In view of low efficiency of digital space regulation and jurisdictional problems of holding accountable for online expression of opinions, all stakeholders should be engaged in solving problems, these are, first, governments, business, civil society and the academic community.

Disdain discourse multiplication in the online mode adversely influences gatherings or people, yet additionally adds to annihilation of deferential and majority rule talk fundamentally. Therefore, the proposed decisions, which may become the subject of further research, should be holistic, globally oriented and responsible.
References


ERBAKAN, V. Turkey, No. 59405/06, ECHR, 6 July 2006.


Recommendation No. R (97) 20 of the Committee of Ministers to member states on “Hate Speech”. Adopted by the Committee of Ministers on 30 October 1997.


