Countering online hate speech: how to adequately protect fundamental rights?

Nave, E.V.R.

2 Legal conceptualization of hate speech

Hate speech, historical or systematic oppressions, and European human rights[12]

ABSTRACT

Today, around 5 billion people communicate through the Internet. While the benefits of online communication are undeniable, we also witness the proliferation of online hate speech, often associated with an increase in offline violence. Internet intermediaries and public bodies have developed frameworks to counter online hate speech. However, current frameworks lack a standardized approach to the conceptualization of hate speech. Some conceptualizations are overbroad, and others are underinclusive; overbroad because they lead to the removal of legal content (e.g. removal tools deleting legal content posted by marginalized communities), and underinclusive as the context of posts by linguistic minorities is often disregarded. This Chapter proposes a new legal conceptualization of hate speech in the European context. It does so by analysing the European regulatory framework through the lens of the first legal conceptualizations of hate speech deriving from critical (race) theory and (black) feminist intersectionality theory. The European focus is justified by the need to standardize at the regional level the legal requirements in current and future policies to counter online hate speech. The methodology is doctrinal, normative, and interdisciplinary legal research. There are two main findings. First, this Chapter suggests that the European regulatory framework needs to explicitly acknowledge the conceptualization of hate speech by critical legal scholars as expressions intended to perpetuate historical or systematic oppression. Second, this Chapter advocates that the conceptualization of hate

1 This Chapter was originally published in the Buffalo Human Rights Law Review 29(2022/2023): 83-145.

2 This Chapter was updated after publication and hence some content deviates from what was previously published. More specifically, references to legal and policy frameworks were updated to reflect the latest available information. Examples include the Council of Europe Committee of Ministers Recommendation CM/Rec(2022)16, the European Union Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act, DSA), the Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence (Artificial Intelligence Act, AI Act), and the Directive of the European Parliament and of the Council on combating violence against women and domestic violence. Cross-references should be read as referring to other references within the present Chapter. The citation style in this Chapter deviates from the standard European citation because it was originally published in an American Law Review.

speech in the European context can only achieve legal cohesion when all European regulatory instruments expressly account for the intersectionality of systems of oppression.

2.1 INTRODUCTION

The Internet enables borderless communications for more than half of the world's population. It connects people who are physically apart and it facilitates the spread of ideas and information. While the benefits of the Internet are undeniable, it also presents a dark side: hateful speech, for instance, tends to spread much faster and farther online,[3] often systematically targeting marginalized groups.[4] As noted by a former Secretary-General of the United Nations, the use of the Internet to promote hateful expressions is one of the most significant human rights challenges arising from technological developments.[5]

Online platforms have been linked to the rise of hate speech and violent conduct. For example, Facebook was accused of contributing to anti-Muslim riots in Sri Lanka and of playing a crucial role by hosting commentary inciting to violence against the Rohingya minority in Myanmar.[6] Other platforms have been associated with mass shootings, e.g. in the cases of Gab in relation to the Pittsburgh synagogue mass shooter and of 8kun with the El Paso killing.[7]

---

3 Binny Mathew et al., *Spread of Hate Speech in Online Social Media*, Proceedings of the 10th ACM Conference on Web Science 173 (2019), https://doi.org/10.1145/3292522.3326034 (last visited Oct 1, 2022).

4 This research acknowledges the ongoing debate about the use of the word "victim" as it may be interpreted to mean that those targeted are in a passive state of subjugation. This research acknowledges also the growing advocacy, especially by members of the civil society, for the use of the word "survivor" as it contains an emphasis on the strength of the people targeted by hate speech. However, it is also debated how this emphasis on strength may burden the targeted community with the obligation to overcome such traumatic experiences. With this background discussion in mind, this research will opt for using the expression "people targeted by hate speech" as much as possible. Nevertheless, the word "victim" may sometimes be used simply for legal coherence as this is the word used in the European Union Victims' Rights Directive 2012/29/EU.

5 U.N. Secretary-General, *Globalization and Its Impact on the Full Enjoyment of All Human Rights"* ¶¶ 26–28, U.N. Doc. A/55/342 (Aug. 31, 2000).

6 Michael Safi, *Sri Lanka Accuses Facebook Over Hate Speech After Deadly Riots*, Guardian (Mar. 14, 2018), https://www.theguardian.com/ world/2018/mar/14/facebook-accused-by-sri-lanka-of-failing-to-control-hate-speech; Alexandra Stevenson, *Facebook Admits It Was Used to Incite Violence in Myanmar*, N.Y. Times (Nov. 6, 2018), https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html.

7 *See* Lizzie Dearden, *Gab: Inside the Social Network Where Alleged Pitts burgh Synagogue Shooter Posted Final Message*, The Independent: Tech (Oct. 28, 2018, 8:10 PM), https://www.independent.co.uk/tech/pittsburgh-synagogue-shooter-gab-robert-bowers-final-posts-online-comments-a8605721.html; Tim Arango, Nicholas Bogel-Burroughs & Katie Benner, *Minutes*

In reaction to these events and due to international pressure, online platforms have started to self-regulate hate speech. However, such self-regulatory efforts often lack a standardized approach to the conceptualization of hate speech that is aligned with human rights. Though some online platforms expressly prohibit hate speech (e.g. Facebook, Twitter, YouTube, LinkedIn, TikTok, Tumblr, Microsoft), they then differ in their definitions. While Facebook defines hate speech as a "direct attack against people – rather than concepts or institutions – on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease,"[8] others do not refer directly to hate speech and focus instead on the prohibition of expressions based on their harmful impact (e.g. Reddit, WhatsApp, LinkedIn).[9]

A specific example of how the platforms' definitions of hate speech may not be aligned with human rights standards is Facebook's definition of "*protected categories*." In 2017, controversies arose when Facebook employed a definition of the categories protected from hate speech, disregarding the protections assigned under international and European human rights law to marginalized groups. In this case, that definition led to the removal of a post suggesting that "*all* white people were racist" but authorized a post incentivizing the "killing of *radicalized* Muslims." This decision was based on the justification that "radicalized Muslims" was a subgroup of a protected marker (i.e. religion), while "all whites" was more generic, thus supposedly more impactful, and therefore deemed more important to protect.[10]

Automated content moderation tools have been said to often be either overbroad or underinclusive. Overbroad because they take down content with no legal basis for removal (online hate speech detection tools have been under scrutiny for racial and queer[11] biases), and underinclusive as they often disregard context or content shared by linguistically marginalized groups.

Before El Paso Killing, Hate Filled Manifesto Appears Online, N.Y. Times (Aug. 3, 2019), https://www.nytimes.com/2019/08/03/us/patrick-crusius-el-paso-shooter-manifesto.html.

8 *Facebook Community Standards: Hate Speech*, META TRANSPARENCY CENTER, https://transparency.fb.com/en-gb/policies/community-standar ds/hate-speech/ (Jan. 13, 2023).

9 *See Content Policy*, REDDIT, https://www.redditinc.com/policies/con tent-policy#:~:text=Abide%20by%20community%20rules.,with%20or%20 disrupt%20Reddit%20communities.&text=Respect%20the%20privacy%20of%20others; *WhatsApp Terms of Service*, WHATS APP, https://www.whats app.com/legal/terms-of-service/?lang=en; *Professional Community* Policies, LINKEDIN, https://www.linkedin.com/legal/professional-community-policies.

10 Julia Angwin, ProPublica & Hannes Grassegger, *Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children*, PROPUBLICA (June 28, 2017, 5:00 AM), https:// www.propublica.org/article/facebook-hate-speech-censorship-internal-do cuments-algorithms.

11 This research uses the term "queer" as an umbrella term to refer to all LGBTIQ+ people, though it acknowledges the ongoing discussion that using the full acronym can be beneficial as a statement to expressly recognize the historically most marginalized groups within the queer community.

To date, there is no legally binding definition of *hate speech* in European or international human rights law. States and public bodies have passed legislation regulating online hate speech but controversies arise on how to conceptualize hate speech and on how to design effective legislation compliant with human rights standards.

The main questions that this Chapter seeks to answer are: (1) what are the main elements of hate speech under European human rights law?, (2) do they align with the original conceptualization of hate speech by critical legal theory?, and (3) to what extent do they require further clarification? By addressing these questions, this Chapter aims to clarify the main aspects of a legal conceptualization of hate speech, grounded in critical legal theory, laying the foundation for an analysis of advances and shortcomings in the European regulatory framework. The focus is on the European context as there is a need to systematize at the regional level the legal requirements for current and future hate speech policies.

The methodology is composed of doctrinal, normative, and interdisciplinary legal research. Doctrinal research focusing on applicable legal frameworks to online hate speech in Europe will contribute to clarifying the existing legal standards. Normative research will identify and address legal loopholes. Interdisciplinary legal research will investigate the interplay between European human rights law and critical legal (race) theory and (black) feminist intersectionality theory. These last two theoretical frameworks were selected as the term (racist) "hate speech" was coined and conceptualized within these fields.

Section 2.2 explores the legal foundations of what hate speech is, what its consequences are, and how it should be regulated from a critical legal perspective. The original legal conceptualization of racist hate speech by critical race theory is key to understanding that hate speech is used against historically and systematically oppressed groups. The insights by critical legal theory also help to understand the impact and harm of hate speech by highlighting the cumulative effects of continued exposure to hate speech and the intersectionality of systems of oppression (race,[12] gender, sexual orientation, etc.). This Section explores the legal foundations of the regulation of hate speech in three different periods: from the Enlightenment, passing by the 1980s and the insights from critical race theory, and to present times. This Section highlights how freedom of expression was never understood as an absolute right and how, since the start of the debate about systematic marginalization, exceptions to free speech have always been accounted for. It concludes with an analysis of the current legal challenges related to hate speech. For instance, the need

12 This research rejects theories of different human "races" as all humans belonging to the same species. However, this research refers to "race" or "racialized" groups as a means to expose a colonial and imperial process whereby a dominant group ascribes to another group a racial identity for the purpose of continued exclusion and domination.

to grant protection to people increasingly targeted by misogynistic and queerphobic hate speech, as well as hate speech targeting people with disabilities. Another current challenge relates to the digitalization of hate speech and how the legal system now needs to account for a faster and further dissemination of hate speech through the internet.

Section 2.3 investigates the theoretical underpinnings of hate speech at the Council of Europe.[13] This Section focuses both on treaty and non-treaty initiatives. The primary treaty is the European Convention of Human Rights (ECtHR), analysed together with relevant case law by the European Court of Human Rights. Other treaties are the Additional Protocol to the Convention on Cybercrime, concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems, the Convention on preventing and combating violence against women and domestic violence, and the Framework Convention for the protection of national minorities. Non-treaty initiatives selected for this analysis include: recommendations and guidelines by the Committee of Ministers; general policy recommendations by the European Commission against Racism and Intolerance (ECRI); outcomes of the European Ministerial Conferences on Mass Media Policy; outcomes of the Council of Europe Conferences of Ministers responsible for media and new communication services; and the Venice Commission Report on the relationship between freedom of expression and freedom of religion.[14] Section 2.3.4.1. analyses the main non-treaty framework which is the Recommendation CM/Rec(2022)16 of the Committee of Ministers to Member States on combating hate speech. This Recommendation draws on the main jurisprudence of the ECtHR on hate speech and is a cornerstone in the clarification of the main elements of hate speech in this Chapter.

Section 2.4 explores the main elements of hate speech in the substantive regulation at the European Union (EU) level. This Section starts by examining the EU's general principles and primary sources such as the Treaty of the EU and the Charter of Fundamental Rights of the EU. It then explains the main advances in the regulation of hate speech in secondary sources of the EU law

13 This Part's structure builds on the work of Tarlach McGonagle, *The Council of Europe Against Online Hate Speech: Conundrums and Challenges*, Council of Europe Conference of Ministers responsible for Media and Information Society "Freedom of Expression and Democracy in the Digital Age: Opportunities, Rights, Responsibilities" 40, 44 (2013).

14 The Parliamentary Assembly of the Council of Europe (PACE), one of the two statutory organs of the Council of Europe, has also been an important actor working to counter hate speech, *e.g.* Resolution 1510 (2006), Recommendation 1805 (2007), Resolution 1743 (2010) and also through the No Hate Parliamentary Alliance. United Nations, Office of the High Commissioner for Human Rights (2011) Relevant Council of Europe Standards and Policies on the Prohibition and Prevention of "Hate Speech", Prepared by Directorate General of Human Rights and Legal Affairs (DGHL), available at <https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/ICCPR/Others2011/CouncilofEurope.pdf> accessed 21 November 2024. Nevertheless, this work prioritizes the analysis of Recommendations by the Committee of Ministers as the decision-making body of the Council of Europe.

such as: the Council Framework Decision on combating certain forms and expression of racism and xenophobia by means of criminal law; the Audiovisual Media Services Directive; resolutions adopted by the European Parliament (EP); the Regulation of the EP and of the Council on a Single Market for Digital Services (Digital Services Act, DSA);[15] the Regulation of the EP and of the Council Laying down harmonized rules on artificial intelligence (Artificial Intelligence Act, AI Act); and Directive (EU) 2024/1385 of the EP and of the Council on combating violence against women and domestic violence.[16] Finally, this Section will explore the EC communication from December 2021 on its intention to extend the list of EU crimes to include hate speech and hate crime. In doing so, it does not focus on the procedural regulation of online hate speech as that pertains to the corporate human rights due diligence responsibilities of internet intermediaries moderating illegal content. Rather, the scope of this Chapter focuses on the substantive conceptualization of hate speech.

Section 2.5 concludes with a summary of the main elements in the legal conceptualization of hate speech rooted in European human rights law and supported by notions of critical theory and intersectionality advanced by the (black) feminism scholarship. Though the main elements in the conceptualization of hate speech were clarified in CM/Rec(2022)16, this Chapter presents two main findings. First, it is critical that the European regulatory framework explicitly acknowledges the scholarship of critical legal scholars in that they conceptualized hate speech as expressions intended to perpetuate historical or systematic oppressions. Second, this Chapter advocates that the conceptualization of hate speech in the European context can only achieve legal cohesion when all European regulatory instruments expressly account for the intersectionality of systems of oppression.

2.2 Legal Theoretical Foundations of Hate Speech

To date, there is no legally binding definition of *hate speech* in European or international human rights law. As noted by McGonagle, hate speech has been a "term of convenience" as it is often used to refer to a wide range of extremely negative content which would otherwise be very difficult to refer to.[17] For

15 The DSA also amends Directive 2000/31/EC, also referred to as the e-Commerce Directive.

16 European Commission, Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401385> accessed 21 November 2024.

17 Tarlach McGonagle, Minority Rights, Freedom of Expression and of the Media: Dynamics and Dilemmas 317 (2011). *See also* Podcast series by Katie Pentney "Decoding Hate," Episode 2 "The Hate You Tweet" with Tarlach McGonagle, 10 February 2021, funded by OSCE Representative on Freedom of the Media, #SAIFE project, available at <https://www.

example, the term "hate speech" has been informally employed to cover a wide range of situations from incitement to violence, hatred, bias, prejudice, insults, or defamation.[18]

This Section explores the legal foundations of hate speech by focusing on the framework developed by critical legal theory, and more particularly, critical race theory (CRT). The framework provided by CRT is a key starting point in this Chapter for the conceptualization of hate speech, as this was the scholarship in the 1990s that coined the term referring to "racist hate speech." CRT is a derivative body of critical legal theory. Critical legal theory, as per Young's definition, is a field of research that analyses society through its historical and sociological contexts.[19] CRT builds on these premises to challenge ahistoricism and to underline how current social and institutional inequalities derive from periods where racist intentions and practices were clearly outspoken.[20] This Section applies critical legal thinking to explain the first legal conceptualization of hate speech, the impact and harm it causes, the foundational legal debates on how to protect fundamental rights, and, current challenges to regulation.

2.2.1 First Legal Conceptualization

The concept of hate speech was coined by the legal scholarship *critical race theory* (CRT) in the early nineties in the United States of America in reference to "racist hate speech."[21] Following a 1990 report by the National Institute Against Prejudice and Violence that highlighted high levels of ethno-violence toward minority students on campuses, many universities and public bodies adopted regulations prohibiting speech stigmatizing racial minorities and other historically subordinated groups. These regulatory initiatives limiting racist expression sparked significant debate in American society and they were met with a strong sentiment of appreciation by members of the communities

decodinghatepod.com/episodes/episode-05-the-anywhere-.w orkout-lhgdz-D0JBu> accessed 4 October 2021; Tarlach McGonagle, *The Council of Europe Against Online Hate Speech: Conundrums and Challenges*, *in* Freedom of Expression and Democracy in the Digital Age: Opportunities, Rights, Responsibilities 40, 44 (2013).

18 James B. Jacobs & Kimberly Potter, Hate Crimes: Criminal Law and Identity Politics 11 (2001).

19 Iris Marion Young, Justice and the Politics of Difference 5 (2011).

20 Mari J. Matsuda et al., Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment 6 (2018); Derrick Bell, And We Are Not Saved: The Elusive Quest for Racial Justice (2008).

21 Matsuda et al. (n 20), 1.

targeted by such hateful speech as well as strong resentment by free speech First Amendment absolutists.[22]

The critical race legal scholars, many being racialized and themselves targets of hate speech, underlined the urgency to halt "racially abusive hate speech" and advocated for the restriction of freedom of expression in cases of racist speech.[23] CRT challenges ahistoricism and underlines how current social and institutional inequalities derive from periods where racist intentions and practices were clearly outspoken.[24] CRT aims to inspire legal and political systems that are informed about the victims' lived experiences.[25] More specifically, this scholarship seeks to provide evidence for how hate speech negatively affects the victim's rights, including dignity, non-discrimination, equality, participation in public life, expression, association, and religion, as well as how hate speech has the potential to inhibit self-fulfilment, self-esteem, and inflict physical harm.[26]

Matsuda, a critical race scholar who helped first conceptualize racist hate speech, contends that racist speech should not be treated as protected discourse under the First Amendment because it is the continuation of subjugation of groups historically oppressed.[27] Matsuda advocates for the prosecution of the worst forms of hate speech to provide public redress for the most serious harm. She proposes three elements to support the identification of the worst forms of racist hate speech: 1) the message is of racial inferiority and all members of the target group are considered alike and inferior; 2) the message is directed against a historically oppressed group and reinforces a historically vertical relationship; 3) the message is persecutory, hateful and degrading.[28] This Chapter explores how Matsuda's conceptualization of hate speech applies to times when digital technology is pervasive and where people are not only being targeted with hate speech on the basis of their race, but also gender, sex, sexual orientation, disabilities, and age.[29]

22 U.S. Const. Amend. I ("Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances").
23 Matsuda et al. (n 20), 2.
24 Matsuda et al. (n 20), 6.
25 Matsuda et al. (n 20), 6; Jean Stefancic & Richard Delgado, Critical race theory: The Cutting Edge (2000); Richard Delgado & Jean Stefancic, Understanding Words That Wound (2019).
26 Matsuda et al. (n 20), 25.
27 *Id.* at 35.
28 *Id.* at 36
29 *Id*. at 16.

2.2.2 Impact and Harm

Along with the progress achieved by CRT on the conceptualization of racist hate speech, CRT was also the first line of scholarship to formally identify the harm caused to people when targeted by hate speech. Delgado and Stefancic highlight physical, psychological, and economic harm caused to people targeted by hate speech.[30] The short-term physical harms vary from rapid breathing, headaches, raised blood pressure, dizziness, and rapid pulse rate. In fact, scientists suspect that the high blood pressure and higher deaths from hypertension, hypertensive disease, and stroke from which many African Americans suffer could be associated with greater racial discrimination.[31] Potential long-term physical harm may in the worst cases lead to hate crime.[32]

With regard to psychological harms, victims of hate speech may experience fear, nightmares, low self-esteem, withdrawal from society (they forego their own right to freedom of expression), post-traumatic stress disorder, psychosis, anger, depression, and rejection of identification with their own race.[33] These effects have a different impact depending on the age of the victim. For example, children are believed to be among the most easily damaged by racial name-calling.[34] This negative impact on youngsters can be heightened when parents experience discriminatory practices themselves and have to put energy into overcoming their own trauma while educating their children.[35] Finally, psychological harms can also lead to self-harmful behaviours.[36]

People targeted by hate speech also experience economic harms. For example, research has shown that racialized students at white-dominated universities may earn lower grades as a result of stress caused by the continuous exposure to racist behaviour.[37] Similarly, racialized people who manage to succeed academically and professionally are often in white-dominated environments and, thus, are more likely to encounter racism. As explained by Matsuda, such experiences may lead hate speech victims to change jobs, forgo education, avoid public spaces and restrict their own freedom of expression to avoid receiving hate messages.[38]

In assessing the harm caused to the targets of hate speech, another key element from CRT that this Chapter explores is the notion of *intersectionality*.

---

30 Delgado & Stefancic (n 25), 12–19.
31 Delgado & Stefancic (n 25), 13.
32 Delgado & Stefancic (n 25), 5.
33 Delgado & Stefancic (n 25), 14.
34 Joe R. Feagin & Debra Van Ausdale, The First R: How Children Learn Race and Racism (2001).
35 Joe R. Feagin & Debra Van Ausdale (n 34).
36 A. Sumner et al., *Association of Online Risk Factors with Subsequent Youth Suicide-Related Behaviors in the US*, JAMA Network Open (Sept. 20, 2021), https://jamanetwork.com/journals/jamanetworkop en/fullarticle/2784337Steven'.
37 Stefancic & Delgado (n 25), 111–121.
38 Matsuda et al. (n 20), 24.

Intersectionality is an analytical framework that explains how elements of someone's social and political identities merge to create different forms of discrimination or privilege. Although the call for attention to intersectionality dates back to Black feminists from the 19th century, when Cooper questioned the overlap of women and race questions,[39] Crenshaw introduced the concept in legal scholarship in 1989.[40] She proposed intersectionality as both a metaphor for crossing categories of discrimination and as a means to show the shortcoming of approaches that seek to isolate systems of oppression.[41] She warned about how the politics and hateful discourse of race and gender have worked to exclude and marginalize especially racialized women.

Crenshaw further suggested the legal concepts of discrimination need to be revised if they are intended to serve as remedies to historical or systematic oppression.[42] Notably, even though the primary intersections explored by the intersectionality theory were race and gender, Crenshaw clarified that the concept could and should be expanded by considering characteristics such as class, sexual orientation, and age.[43] To help clarify the contextual systems of oppression, it is valuable to highlight how Matsuda insists that legal scholars should listen to the victims of hate speech and recognize their entitlement to directly express their concerns and fears; Matsuda claimed that victims would only find redress through such a representation process.[44] This expansive consideration of elements contained in the research of intersectional systems of oppression is essential in the analysis developed further in this Chapter.

Hate speech also harms the perpetrator and society as a whole. For instance, as the perpetrator of hate speech deepens their hateful beliefs, they can develop a paranoid mentality with respect to the community that they routinely denigrate, leading to the spread of hateful beliefs within the community of the perpetrator.[45] Finally, hate speech impacts society altogether as it challenges the fundamental value of equality, and equal respect and dignity, security and the rule of law.[46] In fact, social scientists have shown how people

39 Anna Julia Cooper, A Voice from the South 21 (Zenia, The Aldine Printing House 1892); Anna Carastathis, Intersectionality: Origins, Contestations, Horizons 15 (University of Nebraska Press 2016).

40 Kimberlé Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1989 U. CHI. LEGAL F. 139, 141 (1989).

41 Kimberle Crenshaw, *Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Colour* (1991), Stanford Law Review, Vol. 43, No. 6 (Jul., 1991), pp. 1241-1299.

42 Crenshaw (n 40), 140.

43 Crenshaw (n 41), 1244-45.

44 Matsuda et al. (n 20), 114.

45 Matsuda et al. (n 20), 24.

46 Jeremy Waldron, The Harm in Hate Speech 92 (2012).

targeted by hate crimes take much longer to recover compared to people targeted by crimes without racist motivation.[47]

Finally, this Chapter agrees with critical race theorists in that it is important to consider the cumulative effect of continued exposure to hate speech. More specifically, groups of people who have been historically targeted by hate speech may develop emotions of self-hatred, especially when the victim internalizes the negative perception of who they are.[48] The various *types* of harm caused by hate speech are real and require legal redress to avoid perpetuating historical oppressions. As advocated by Parekh, hate speech should be restricted both "for what it is and for what it does."[49] In establishing a legal framework for countering hate speech, it is important to understand how the regulation of hate speech interplays with other rights and to question if there are different *degrees* of hate speech requiring different legal courses of action.

2.2.3 Regulation and Balancing Conflicting Rights

In the previous subsections 2.2.1. and 2.2.2, this Chapter explained the original conceptualization of (racist) hate speech and its impacts as presented by critical race theory. The following paragraphs explore the foundational debates on the regulation of hate speech; more specifically, the debates on the balance of competing rights when regulating hateful expression. As any form of regulation of hate speech inevitably affects the exercise of the right to freedom of expression, this subsection explores how legal scholars have balanced competing rights – freedom of speech versus dignity[50] and prohibition of discrimination.[51] This analysis will focus on two different periods: first, the Age of Enlightenment, and second, the period from the 1990s marking the first conceptualization of racist hate speech by critical race theory.

Analysing the origins of the right to free speech and its original conceptualization during the Age of Enlightenment, it is possible to conclude that, at its conception, free speech was not considered an absolute right in situations where it inflicted *serious harm* to others. The right to freedom of expression

47 Frederick M. Lawrence, *The Punishment of Hate: Toward a Normative Theory of Bias-Motivated Crimes*, 93 MICH. L. REV. 320, 342–343 (1994).

48 Richard Delgado, *Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C. R.-C. L. L. REV. 133, 137 (1982); Charles R. Lawrence, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 Stan. L. Rev. 317, 201 (1987).

49 Bhikhu Parekh, *Is There a Case for Banning Hate Speech?* in The Content and Context of Hate Speech: Rethinking Regulation and Responses 37 (Michael Herz & Peter Molnar, eds., 2012).

50 The right to human dignity is contained in Art. 1 of the EU Charter of Fundamental Rights.

51 The right to freedom from harm is herewith used to refer broadly to the right to dignity and the right to non-discrimination.

was a prominent debate in the Liberalist ideals from the 17th and 18th centuries.[52] Liberalism prioritizes individual freedoms such as freedom of expression, thought, and association. This period was marked by an emphasis on the importance of preserving the marketplace of ideas, in which the truth would only emerge from a free, transparent and public competition of ideas.[53] However, even the most renowned liberal thinkers did not advocate for an absolutist conception of freedom of expression. In fact, Locke proposed a non-absolutist notion of tolerance when he defended opinions contrary to human society, or the moral rules necessary to the preservation of society, should not be legally protected.[54] Mill went further to propose legal restrictions on freedom of expression particularly when it caused "harm to others."[55]

Despite the existence of some considerations of the non-absolutist notion of freedom of speech at the beginning of Liberalism, it was in the 1990s that the legal debate about historical oppressions and the restriction of racist hate speech ignited in the United States. There are two key elements in the debate on the regulation of hate speech from the beginning of critical race theory to the present times. First, there were new stronger arguments for the protection of marginalized communities which were challenged by traditional constitutionalist US scholars. Second, there was more formal discussion about the harm in hate speech as the continuation of historical oppression.

First, critical legal scholars challenged the Liberalist conception of equal opportunities to exercise rights and underlined the need to restrict hateful speech, as it perpetuated harmful practices toward the already marginalized community.[56] As noted by Waldron, perpetrators of hate and degradation intend to undermine dignity and destabilize the identity attributes shared by members of targeted oppressed communities.[57] Also, as proposed by MacKinnon, "a well-ordered society" must assure that members of marginalized groups live in a dignified manner.[58]

However, they were met with strong opposition by other scholars from the United States Constitutional tradition. For example, Dworkin stated that no individual should be shielded from content that may negatively impact

52 Liberalism was a societal system that placed the emphasis on individual liberty, equality, and consent to be governed, and supported access to common resources in proportion to the individual's contribution. Liberal ideals replaced feudalism, which established a static set of social classes with different fixed tasks and different fixed possibilities to access resources, indifferent to an individual's contribution to the common good.

53 Caitlin Ring Carlson, Hate Speech 10 (2021).

54 John Locke, A Letter Concerning Toleration (Merchant Books 2011) (1765).

55 John Stuart Mill, On Liberty and Other Essays (John Gray ed., Oxford Univ. Press 1998) (1859).

56 *See Section* 2.1 and 2.2.

57 Waldron (n 46), 92.

58 Catharine A. MacKinnon, Only words (1993).

their self-esteem.[59] Scanlon emphasized stronger protection for the speaker, pluralism and non-interference, often at the expense of the rights of the group targeted by that hate speech, despite admitting that such speech can cause harm.[60] Traditional American scholars' resistance to the regulation of hate speech is flawed because there were already many situations limiting freedom of expression,[61] and such an absolutist interpretation of the First Amendment's response to hate speech simply perpetuates racism.[62] Additionally, these traditional constitutionalist American scholars fail to recognize the limits in their positionality,[63] as they fail to recognize their position of power and privilege and their likely (un)conscious bias. As advocated by Cohen-Almagor, the American viewpoint of neutrality defended by Scanlon is inherently false for it fails to recognize the inequalities and discriminatory practices in access and exercise of basic freedoms.[64]

Second, hate speech is a term that has been used informally across many fields such as law, sociology, criminology, and psychology, to refer to a vast number of harmful expressions including incitement to hatred, insults, defamation, bias, and prejudice. All these expressions could happen simultaneously or separately and could respectively cause different degrees of harm. This Chapter aligns with Post's claim that for an expression to be hate speech it needs to be "extreme" in nature because no legal order can aim to abolish emotions of intolerance and dislike.[65] However, it is herewith proposed a broader interpretation of extreme because hate speech is in fact, in and of itself, harmful expression. There is nevertheless a distinction to be made between hate speech and the most serious forms of hate speech which should have

59 Ronald Myles Dworkin, Freedom's Law: The Moral Reading of the American Constitution 260 (1996).

60 See Thomas Scanlon, *Freedom of Expression and Categories of Expression Principles of Expression and Restriction: A First Amendment Symposium*, 40 U. PITT. L. REV. 519, 527 (1979). Nevertheless, it should be noted that, in a more recent work published in 2018, Scanlon acknowledged that insult and harassment causing psychological harm might be grounds for exclusion of expression, *e.g.* Thomas Scanlon (2018) A Framework for Thinking about Freedom of Speech and some of its Implications, available at <https://www.law.berkeley.edu/wp-content/uploads/2018/10/Freedom-of-Speech-Berkeley.pdf> accessed 21 November 2024. See *also* Raphael Cohen-Almagor, *Racism and Hate Speech – A Critique of Scanlon's Contractual Theory*, 53 FIRST AMENDMENT STUDIES 1, 2 (2019).

61 Examples include privacy, individual reputation, protection of intellectual property, regulation of economic markets, speech infringing public order. *See* Delgado & Stefancic (n 25), 11, 34; Frederick Schauer, *Categories and the First Amendment: A Play in Three Acts*, 34 Vand. L. Rev. 265, 270 (1981).

62 Schauer (n 61), 270; Parekh (n 49), 222.

63 Positionality is the concept that someone's personal experiences of race, gender, class, etc., and location in time and space affect one's view of the world. Issues of positionality refute ideas of neutrality, and objective research shows that any researcher will inevitably design and conduct research in a subjective way influenced by their positionality in the world.

64 Cohen-Almagor (n 60), 22.

65 McGonagle (n 17), 14 (citing Ivan Hare & James Weinstein, Extreme Speech and Democracy 123 (2009)).

implication in the actionable legal area i.e. civil and administrative law for hate speech and criminal law for the most serious forms of hate speech.

In assessing whether an expression is extreme and causes serious harm, States should investigate the context in which the expression was manifested and should specifically investigate if the targeted community has been historically oppressed. In examining if the targets of hate speech have been historically oppressed, as pointed out by Boyle and Baldaccini, it is important to investigate the "core mischiefs" of hate speech i.e. the impact of hate speech on the exercise of other rights.[66] Examples of this exercise would be to look for how the targets of hate speech have been accessing among others their rights to dignity, to have a house, to have a job, and to education by investigating the activities and claims by civil rights movements.

2.2.4 Current Legal Challenges

The critical legal theory, foundational in the conceptualization of hate speech explained in Section2.1, has been particularly challenged by various recent developments in terms of impact as well as mediums, and reach. First, the impact of hate speech is now broader. Recent data shows that, aside from the racialized community and from women, hate speech now also affects more seriously LGBTIQ+ people, people with disabilities, and members of religious groups.[67] The expanded recognition of the groups targeted by hate speech calls for a more formal legal acknowledgment of the intersectionality aspects of the victims and the present regulation does not currently reflect such development, often still being solely applicable for cases of racist hate speech.

Second, there are also new regulatory challenges related to the mediums and reach of hate speech since the beginning of the digitalization era. With the advent of the Internet, hate speech is spreading faster and further online,[68] potentially also leading to an increase in offline violence.[69] Additionally, the online spaces where hate speech has been spreading are mostly run by U.S.-based companies operating on the basis of American traditional constitutionalism, i.e., the First Amendment which predicates on an almost absolutist interpretation of freedom of expression. The current dominance of U.S.-based companies running online platforms operating worldwide has created difficulties

66 *Id.* at 319 (citing Kevin Boyle & Anneliese Baldaccini, A Critical Evaluation of International Human Rights Approaches to Racism 152 (2017)).

67 *The EU Code of Conduct on Countering Illegal Hate Speech Online*, European Commission, https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en (last visited March 12, 2023).

68 Mathew et al. (n 3).

69 Karsten Müller & Carlo Schwarz, *Fanning the Flames of Hate: Social Media and Hate Crime*, 19 J. EUR. ECON. ASS'N 2131 (2021).

in a decentralized public and for contextual oversight of the hate speech rules guiding our freedom of expression and right to non-discrimination online.

Third, the fast spread of harmful content online has led to the use of the term "hate speech" by various scholarships in a much broader way than in legal scholarship. At present times, hate speech continues to be a term *informally* used across many research fields to cover a spectrum of expressions including: (a) criminal offenses; (b) problematic expressions, which although not criminal offenses, could be prohibited under civil or administrative law; and, (c) content which bears no legal implications but still raises issues of respect and tolerance.[70]

The following Sections 2.3 and 2.4 will first analyse respectively how hate speech is conceptualized at the Council of Europe and at the European Union and seek to clarify how these regulatory systems align with the critical legal foundation discussed in Section 2.2. They will also address the current regulatory challenges with respect to online hate speech as explained in this subsection.

2.3 Approaches to Hate Speech at the Council of Europe

2.3.1 General Objectives

There is no explicit mention of hate speech in the European Convention for the Protection of Human Rights and Fundamental Freedom (ECHR) or in any other treaties of the Council of Europe (the CoE or The Council). Still, the CoE developed various instruments which are useful in the regulation of hate speech. All these instruments follow the main objectives of the Council which are to uphold human rights, democracy, tolerance, non-discrimination and the rule of law.[71]

Nevertheless, despite this alignment in terms of general legal principles, the various regulatory initiatives by the Council to address hate speech pursue different specific goals and contain different procedural and substantive thresholds. As a result, there is a significant variety of strategies ranging from legal action countering hate speech, to education, and promotion of increased representation of minorities in the media.

---

70 U.N. High Commissioner for Human Rights, Report of the United Nations High Commissioner for Human Rights on the Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred, ¶12, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013).

71 Statute of the Council of Europe preamble, May 5, 1949, 87 U.N.T.S. 103.

2.3.2 ECHR and ECtHR Jurisprudence

This Section focuses on European human rights law governing hate speech by addressing, firstly, the most relevant provisions in the European Convention of Human Rights (ECHR) and subsequently, investigating the jurisprudence of the European Court of Human Rights (ECtHR) in the application of such provisions.

The ECHR does not contain a direct reference to hate speech. Instead, it regulates other rights impacting the regulation of hate speech. These rights are often in opposition and their application requires legal reasoning to understand how to draw the balance between competing human rights.

From the perspective of the speaker of potentially hateful content, the most relevant article is the right to freedom of expression and to what extent there can be restrictions on its exercise (Art. 10(2)). From the perspective of the people targeted by potentially hateful speech, the most relevant provision is the prohibition of abuse of rights (Art. 17). Additionally, the ECHR also prescribes the prohibition of discrimination (Art. 14), later expanded in Protocol 12,[72] and grants the right to respect for private life (Art. 8). To fully grasp the legal reasoning to strike the balance between competing rights, it is important to look into the jurisprudence of the ECtHR.

The first reference of the ECtHR to hate speech dates from 1999,[73] though the term was not developed at the time. Still, and despite the fact that the Court never provided a concrete and fixed definition of hate speech, it developed the meaning of hate speech in various instances thereafter. The ECtHR has generally declared that statements attacking or casting in a negative way an entire group on the basis of for example ethnicity, race, and religion, are in contradiction with the underlying values of tolerance, social peace and non-discrimination prescribed by the ECHR.[74] Yet, as the conceptualization of the harm caused by hate speech continues to be very context-dependent, it is crucial to clarify the legal strategies deployed by the Court in cases potentially amounting to hate speech to understand the applicable European human rights standards.

Before explaining the legal strategies in more detail, it is important to point out some of the key interpretative principles guiding the Court in the effort

72 Protocol 12 to the European Convention on Human Rights art. 1, Nov. 4, 2000, E.T.S. 177.

73 Drawing on the work of McGonagle, the term hate speech was first used in four judgments of the ECtHR, all of July, 8, 1999: Sürek v. Turkey (No. 1), App. No. 26682/95, ¶ 62 (July 8, 1999), https://hudoc.echr.coe.int/eng?i=001-58279; Sürek & Özdemir v. Turkey, App. Nos. 23927/94 & 24277/94, ¶ 63 (July 8, 1999), https://hudoc.echr.coe.int/eng?i=001-58278; Sürek v. Turkey (No. 4), App. No. 24762/94, ¶ 60 (July 8, 1999), https://hudoc.echr.coe.int/eng?i=001-58298; Erdogdu & Ince v. Turkey, App Nos. 25067/94 & 25068/94, ¶ 54 (July 8, 1999), https://hudoc.echr.coe.int/eng?i=001-58275; McGonagle (n 17), 11.

74 Perinçek v. Switzerland, App. No. 27510/08, ¶ 206 (Oct. 15, 2015), https://hudoc.echr.coe.int/eng?i=001-158235.

to balance competing rights in the ECHR, especially when balancing the right to freedom of expression (Art. 10) and the right to non-discrimination (Art. 14) or right to private life (Art. 8).[75] First, as freedom of expression represents such a cornerstone of the protection of democracy and the rule of law, the Court decides on a case-by-case basis regarding cases on restrictions of freedom of expression.

Second, the Court follows the margin of appreciation doctrine[76] whereby it assigns a certain discretion to national courts in the domestic interpretation and application of ECHR provisions, subject to the Court's supervisory jurisdiction role. The Court has considered that States are better placed to appreciate the meaning of public morals, decency and religion as these may vary considerably from country to country. Thus, instead of having a cross-cutting understanding of the meaning of such concepts, in reviewing domestic cases impacted by such considerations, the Court has sought to assess whether the justifications by national authorities are relevant and sufficient.

Third, the Court defends that the rights prescribed in the ECHR must be "practical and effective."[77] This means that rights in the ECHR must not be interpreted in an elusive and hypothetical way and, as a consequence, the violation of rights in the ECHR should lead to effective remedial procedures (Art. 13).

Fourth, the Court considers the ECHR a "living instrument" which indicates that the rights contained in the ECHR "must be interpreted in the light of present-day conditions."[78] As the living conditions change, rights can too be challenged by unforeseen circumstances and, as such, the ECHR must continue to be interpreted in such a way that continues to grant protection of rights even if applied in different future contexts unanticipated at the time of the drafting.

Finally, the Court developed a positive obligations' doctrine whereby it requires that States ensure every person can exercise the rights in the ECHR

75 E.g. in Kaboğlu v. Turkey, the Court found that there had been a violation of Article 8 emphasizing that the negative impact of the verbal attacks and threats of physical harm against the applicants. Kaboğlu v. Turkey, App. Nos. 1759/08, 50766/10 and 50782/10 (October 30, 2018). In Beizaras and Levickas v. Lithuania, the Court found that there had been a violation of Article 14 in conjunction with Article 8, concluding that the applicants had been discriminated on the grounds of their sexual orientation. Beizaras and Levickas v. Lithuania, Case Number 41288/15 (January 14, 2020). In a similar case, the Court found that the Romanian authorities had failed to discharge their positive obligation to investigate if a verbal homophobic abuse had amounted to a criminal offence. In this case, the Court found a violation of both Articles 14 and 8; see Association ACCEPT and Others v. Romania, Application no. 19237/16 (June 1, 2021).

76 The margin of appreciation doctrine was included in the Preamble to the ECHR with the adoption of the Convention Amending Protocol No. 15 in August 2021.

77 Airey v. Ireland, App. No. 6289/73, ¶ 24 (Oct. 9, 1979).

78 Tyrer v. the United Kingdom, App. No. 5856/72, ¶ 31 (April 25, 1978); Matthews v. the United Kingdom, App. No. 24833/94, ¶ 39 (Feb. 18, 1999).

not only through a non-interference principle but also when necessary through interfering and measures protecting the exercise of rights.[79]

Turning to the legal strategies applied by the Court when ruling on cases concerning hate speech, as explained by former Judge Tulkens, these can be summarized in two possible approaches.[80] In the first approach (explained in Section 2.3.2.1), the ECtHR can apply Art. 17 (prohibition of abuse of rights) and exclude protection of said hateful expression from the ECHR because such conduct violates or limits (to an extent greater than the one provided in the ECHR) any right in the ECHR, and it is therefore considered an abuse of rights. In the second approach (explained in Section 2.3.2.2), the ECtHR can apply Art. 10(2) when the hateful expression is not considered to violate or limit fundamental rights in the ECHR, but it could still amount to a hateful expression that should be restricted if it meets the conditions in Art. 10(2) (explained in more detail in Section 2.3.2.2).

2.3.2.1 Hate Speech as a Clear Abuse of Rights

As per the jurisprudence of the ECtHR in its first approach to hate speech, hate speech can trigger the prohibition of abuse of rights provision (Art. 17 ECHR). Before elaborating on the substance of the case law related to the application of Art. 17 to hate speech cases, a note is necessary on the procedural consequences of the application of the prohibition of abuse of rights.

The ECtHR uses Art. 17 mainly when it considers whether there is a clear abuse of rights because the act *in casu* either violates or limits (further than what is allowed in the ECHR) rights in the ECHR. In such cases, the Court deems the application of a case inadmissible on its merits. Such declaration of inadmissibility means that, as the case represents such a serious abuse of rights, the Court refuses to proceed with the process of balancing of rights and will not proceed to the judgment on the substance. In essence, when Art. 17 is invoked for hate speech cases, it means that these are the most blatant cases of hate speech. Therefore, the Court dismisses such cases as manifestly ill-founded[81] and does not even proceed to the assessment of Art. 10(2) on whether the restriction to the right to freedom of expression of the speaker of hate speech was rightfully applied.[82]

Elaborating on the case law related to the substantive application of the prohibition of abuse of rights (Art. 17), i.e., to the conceptualization of the most

79 Özgür Gündem v. Turkey, App. No. 23144/93, ¶ 43 (March 16, 2000), https://hudoc.echr.coe.int/eng?i=001-58508.

80 Françoise Tulkens, The Hate Factor in Political Speech: Where Do Responsibilities Lie?, Report of the Council of Europe Conference, Warsaw, Sept. 18-19, 2013.

81 ECHR Art. 35(3).

82 Seurot v. France, App. No. 57383/00 (May 18, 2004), https://hudoc.echr.coe.int/fre?i=002-4404. *See also* Factsheet – Hate Speech, September 2020, Press Unit, European Court of Human Rights, 1-5.

serious cases of hate speech, the Court has ruled that these include: negationist and revisionism; incitement to discrimination, hatred or violence on the grounds of race, ethnicity, religion or sexual orientation likely to give rise to feelings of rejection and hostility; threats to democracy by expressions inspired by totalitarian views; and support to terrorism. Each of these forms of the most serious cases of hate speech will now be explained.

First, negationist and revisionist practices negating the Holocaust or other genocides[83] or other crimes against humanities should be considered hate speech and a clear abuse of rights under Article 17. The Court has supported that such expressions amounted to some of the most serious examples of hate speech because they denied clearly established historical facts that cannot be considered historical research and cannot be considered an endeavour pursuing the truth; had the purpose of rehabilitating the National Socialist regime; double victimized the victims accusing them of falsifying history.[84]

Second, acts that incite discrimination, hatred or violence and are likely to give rise to feelings of rejection and hostility have also been considered by the ECtHR as amounting to the most serious cases of hate speech and deemed a clear abuse of rights under Article 17. For example, in Le Pen v. France,[85] the Court agreed with the national court's conviction of incitement to discrimination, hatred and violence toward a group of people because of their origin or membership of non-membership of a specific ethnic group, nation, race or religion. The ECtHR declared the comments made by the president of the French National Front in an interview saying that "the day that there are no longer 5 million but 25 million Muslims in France, they will be in charge" indeed represented Muslim community as a whole in a disturbing light, likely to give rise to feeling of rejection and hostility, and therefore, interference had been necessary to protect the democratic values.

In assessing such cases of hate speech inciting discrimination, hatred or violence, this Chapter suggests the interpretation of the protected categories should be open-ended. Even though the ECtHR has formally recognized the

83 It is noteworthy that the First Additional Protocol to the Cybercrime Convention is the first international treaty that applies to genocides other than the Holocaust (Art 6).

84 Garaudy v. France, App. No. 65831/01, ¶ 23 (March 25, 2003), https://hudoc.echr.coe.int/fre?i=002-4830. Other examples include M'bala M'bala v. France, App. No. 25239/13 (Oct. 20, 2015), https://hudoc.echr.c oe.int/eng?i=001-160358; Honsik v. Austria, App. No. 25062/94 (Oct. 18, 1995), https://hudoc.echr.coe.int/eng?i=001-2362; Marais v. France, App. No. 31159/96 (June 24, 1996), https://hudoc.echr.coe.int/eng?i=001-88275; Williamson v. Germany, App. No. 64496/17 (Jan. 8, 2019), https://hudoc.echr.coe.int/eng?i=001-189777; Pastors v. Germany, App. No. 55225/14 (Jan. 3, 2020), https://hudoc.echr.coe.int/eng?i=001-196148; Peta Deutschland v. Germany, App. No. 43481/09 (Mar. 18, 2013), https://hudoc.echr.coe.int/eng?i=001-114273; Perinçek v. Switzerland, App. No. 27510/08 (Oct. 15, 2015), https://hudoc.echr.coe.int/eng?i=001-139724.

85 Le Pen v. France, App. No. 18788/09, (Apr 20, 2010).

protection against hate speech discriminating only on the basis of the race,[86] religion,[87] and ethnicity,[88] other grounds should also be regarded as impermissible grounds. In fact, the Court recalled in Lilliendahl v. Iceland that discrimination based on sexual orientation is as serious as discrimination based on "race, origin or colour" and underlined the importance to grant protection from hateful and discriminatory speech to groups on the basis of gender and sexual minorities and stressed the historic marginalization that these groups have endured.[89]

An important element in the application of this category of hate speech (incitement to discrimination, hatred or violence on the grounds of race, ethnicity, religion or sexual orientation likely to give rise to feelings of rejection and hostility) is that it requires a negative generalization of the targeted group. For example, in Norwood v. the UK, the ECtHR specified that linking a (religious) group as a whole with a grave act of terrorism was incompatible with the democratic values of tolerance, social peace and non-discrimination.[90]

---

86 *See, e.g.*, Glimmerveen & Hagenbeek v. Netherlands, App. No. 8348/78 (Oct. 11, 1979); Medya FM Reha Radvo ve Iletisim Hizmetleri A. S. v. Turkey, App. No. 32842/02 (Nov. 14, 2006); Simunic v. Croatia, App. No. 20373/17, (Jan 22, 2019).

87 *See, e.g.*, Belkacem v. Belgium, App. No. 343667/14, ¶ 21 (Jun. 27, 20 17), https://hudoc.echr.coe.int/fre?i=001-175941; Gunduz v. Turkey, Ap p. No. 35071/91, ¶18 (Nov. 13, 2020), https://hudoc.echr.coe.int/fre?i=001-6 1522.

88 *See e.g.*, Ivanov v. Russia, App. No. 35222/04, ¶ 10 (Feb. 20, 2007); W.P. & Others v. Poland, App. No. 42264/98, ¶ 47 (Sept. 2, 2004), https://hudoc.echr.coe.int/fre?i=001-66711.

89 Lilliendahl v. Iceland, App. No. 29297/18, ¶ 45 (May 12, 2020) (Nevertheless, this is a controversial case, and some remarks are necessary. On the one hand, this is a remarkable case as the Court confirmed the protection of the queer community from hate speech and, more specifically, the Court sided with the domestic court's view that the homophobic online comment amounted to hate speech. On the other hand, though the Court declared the case inadmissible and manifestly ill-founded, it did not consider the comment as one of the most serious forms of hate speech. The Court stopped short of granting more solid protection to the queer community when it declared that the applicant's comment did not amount to the most serious forms of hate speech claiming that it was not clear the aim was to incite violence and hatred or to destroy rights in the ECHR. This position fails to align with the criteria used to investigate the most serious cases of hate speech. Establishing a parallel with the Le Pen v. France case, some key elements in the comment should have been sufficient for the EctHR to recognize incitement to discrimination and hatred in a disturbing way likely to give rise to a feeling of rejection and hostility, and therefore constitute an example of the most serious cases of hate speech. These include the fact that the comment clearly stated that gay people had a sexual deviation, such said sexual deviation was characterized by derogatory words mainly used to describe animals, and finally the fact that the applicant considered homosexual people disgusting and repulsed their representation in the media. These statements seriously undermine the broader reach of the impact of hate speech online and compromise the right to respect for private life, non-discrimination, and to freedom of information, as they specially legitimize oppression of a historically repressed group. The Lilliendahl v. Iceland case is essential to clarify that homophobic speech is hate speech but cannot be considered as a deviation from the jurisprudence regarding the criteria used to investigate the most serious cases of hate speech.)

90 Norwood v. United Kingdom, App. No. 23131/03, ¶ 13 (Nov 16, 2004).

Similarly, in Pavel Ivanov v. Russia, the ECtHR also declared that "accusing an entire ethnic group of plotting a conspiracy" was a general, vehement attack on one ethnic group and is directed against the Convention's underlying values of tolerance, social peace and non-discrimination.[91]

In assessing the prohibition of hate speech in the form of incitement to discrimination, hatred or violence on the grounds of religion, the Court specified that only cases of criminal offenses should be included under Article 17.[92] More specifically, in Mariya Alekhina and Others v. Russia, the ECtHR ruled that insults to religious beliefs and blasphemy should not be subject to criminal sanctions.[93] This represents a narrower affordance of domestic autonomy within the margin of appreciation doctrine.

Third, another category of hate speech considered serious enough to amount to a prohibition of abuse of rights is speech that threatens the democratic order because it is inspired by totalitarian views. As a rule, the ECtHR will declare inadmissible applications inspired by totalitarian doctrines or which express ideas that represent a threat to the democratic order and are liable to lead to restoration of a totalitarian regime.[94]

Fourth, the Court has also interpreted expressions supporting terrorist activities as serious cases of hate speech under Article 17. In Roj TV A/S v. Denmark, the media service had applied against its conviction for terrorism offenses by Danish courts for promoting the Kurdistan Workers' Party (PKK) through television programs. The ECtHR found that the PKK could be considered a terrorist organization within the meaning of the Danish Penal Code and supported the domestic conviction in light of the margin of appreciation doctrine. However, this case could be read as to dismiss an important debate on the lack of due process in the European Union for the classification of the PKK as a terrorist organization.[95] Therefore, questions arise in this case as to whether the Court conducted an effective investigation to protect media independence and access to information. Moreover, given the ECtHR's interference with the margin of appreciation doctrine in the case Mariya Alekhina

91 *See e.g.*, Ivanov v Russia, App. No. 35222/04, ¶ 20.

92 *Id.* This position confers a better alignment of the ECtHR with international human rights standards on the balance exercise between the restrictions to freedom of expression and freedom of religion.

93 *See, e.g.*, Alekhina v. Russia, App. No. 22519/02, ¶ 224 (July 13, 2006), https://hudoc.echr.coe.int/fre?i=001-73321.

94 *See, e.g.*, Communist Party of Germany v. Federal Republic of Germany, App. No. 250/57 (July 20, 1957), https://hudoc.echr.coe.int/eng?i=001-110191; B.H., M.W., H.P. & G.K. v. Austria, App. No. 12774/87 (Oct. 12, 1989), https://hudoc.echr.coe.int/eng?i=001-1039; Nachtmann v. Austria, App. No. 36773/97 (Sept. 9, 1998), https://hudoc.echr.coe.int/eng?i=001-4399; Schimanek v. Austria, App. No. 32307 (Feb. 1, 2000), https://hudoc.echr.coe.int/eng?i=001-24075

95 *European Court: Decisions Placing the PKK on the List of Terrorist Organizations Annulled*, PRAKKEN D'OLIVEIRA (Nov. 15, 2018), https://www.prakkendoliveira.nl/en/news/2018/european-court-decisions-placing-the-pkk-on-the-list-of-terrorist-organizations-annulled.

and Others v. Russia in the same year as Roj TV A/S v. Denmark, it could be argued that also in the latter case the Court could have at least acknowledged the controversies on the classification of PKK as a terrorist group.

Recapitulating, the application of Article 17 leads the Court to dismiss the case as manifestly ill-founded, thus in principle not even leading to an assessment of the balance of rights. Nevertheless, there were three cases in which the Court, despite considering the application inadmissible, still provided further details as to why specific cases of hate speech were amongst the most serious hateful expressions. First, in Šimunić v. Croatia,[96] a case of offline hate speech, the ECtHR sided with the domestic courts stressing the need to tackle racism and totalitarian ideas shared by prominent sports figures.

Second, in Smajić v. Bosnia and Herzegovina,[97] a case regarding online hate speech, the Court found the domestic courts had rightfully convicted the applicant for national, racial, and religious hatred, discord or intolerance following posts on an internet forum describing military action that could be undertaken against Serb villages. The ECtHR ruled that the penalties of a suspended sentence and a seized laptop had not been excessive and dismissed the claim as manifestly ill-founded.

Third, in Nix v. Germany,[98] also a case regarding online hate speech, the Court sided with the domestic court's decision to consider the applicant criminally liable for an online post of a picture of a Nazi leader and a swastika because it had not been clear the applicant's intent to reject the Nazi ideology, finding the application manifestly ill-founded. In this case, the Court reiterated that Article 10 of ECHR applied to the Internet, as did the conditions for the restriction of freedom of expression in Article 10(2). The ECtHR added that States which have experienced the Nazi horrors may be regarded as having a special moral responsibility to distance themselves from the Nazi ideology in light of their historical role and experience from the mass atrocities perpetrated by the Nazis.[99] This case could be read as to assign special moral responsibilities of States with history of totalitarian regimes to counter online hate speech.

In summary, the Court applies Article 17 in cases of criminal hate speech, which in its view should only be applied to the most serious cases of hate speech. When analysing the jurisprudence of the ECtHR on Article 17 in the light of the original conceptualization by Matsuda[100] of hate speech as ex-

96 Šimunić v. Croatia, App. No. 20373/17 (Jan. 22, 2019), https://hudoc.echr.coe.int/eng?i=001-189769.

97 Smajić v. Bosnia & Herzegovina, App. No. 48657/16 (Jan. 18, 2018), https://hudoc.echr.coe.int/eng?i=001-180956.

98 Nix v. Germany, App. No. 38285/16 (Mar. 13, 2018), https://hudoc.echr.coe.int/eng?i=001-182241.

99 *Id.* ¶ 47.

100 The conceptualization of (racist) hate speech introduced by Matsuda is explained on Section 2.2.1. of this Chapter.

pressions perpetuating historical or systematic oppressions, it is possible to conclude the Court fails to consistently mention historical or systematic oppressions as a key element of hate speech and fails equally to recognize the intersectionality of systems of oppression perpetuated by hate speech.

2.3.2.2 No Clear Abuse But Hate Speech is Prohibited

Under the second approach, i.e., in cases where there is no clear abuse of rights as per Article 17 ECHR and where the application is not considered inadmissible on its merits, the ECtHR applies the right to freedom of expression (Art. 10).[101] An essential point of departure is that the Court posited that freedom of expression applies "not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population."[102]

Any restriction on the right to freedom of expression speech needs to follow the criteria contained in Article 10(2). That is, the Court considers a restriction of freedom of expression to be legal if it is: (i) prescribed by law; (ii) in pursuit of one or more specified legitimate interests (national security, territorial integrity or public safety, prevention of disorder or crime, for the protection of health or morals, reputation or rights of others, prevention of the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary); and (iii) necessary in a democratic society.

When assessing the necessity criterion, the ECtHR evaluates whether there is a pressing social need, whether the restriction is proportional and the relevance and sufficiency of States' justifications. It is relevant to highlight that for restrictions to be deemed necessary, they need to aim at protecting the interests of national security, public safety, the prevention of disorder or crime, the protection of health or morals and the protection of the rights and freedoms of others.[103] In this evaluation, the ECtHR leaves to the States a margin of appreciation, recognizing the diversity in social and legal traditions across all Member States.[104]

---

101 It is noteworthy that although Article 10 (right to freedom of expression) has been the main point of departure in the ECtHR jurisprudence, hate speech can impact a variety of other rights such as non-discrimination, life, association, etc.

102 Handyside v. UK, App. No. 5493/72, ¶ 49 (Dec. 7, 1976), https://hudoc.echr.coe.int/eng?i=001-57499.

103 For a detail study, see *e.g.* Janneke Gerads (2013) "How to improve the necessity test of the European Court of Human Rights." *International journal of constitutional law* 11.2: 466-490.

104 The margin of appreciation doctrine was used first in Lawless v. Ireland, E.Ct.H.R. Ser.B, 1960-61, para.90, p.82. Nevertheless, this doctrine's prominence was more definitively affirmed in Handyside v. U.K. in 1976. For a detailed analysis see Hutchinson, M. R. (1999) The Margin of Appreciation Doctrine in the European Court of Human Rights. *The International and comparative law quarterly*. [Online] 48 (3), 638–650.

When assessing the severity of the hateful expression, the ECtHR has developed a set of standards which Rosenfeld describes as the "contextual variables"[105] approach. Generally, this approach has considered the victims' perspectives;[106] political and social background;[107] the intent of the speaker;[108] the speaker's status or role in society;[109] the content of the expression;[110] the reach of the expression;[111] and, the nature of the audience.[112] With regard to the element of intentionality, this Chapter aligns with the dissenting opinion by Judges Ryssdal, Bernhardt, Spielman and Loizou in Jersild v. Denmark in that the presumably good intentions of the person disseminating hate speech are not enough to dismiss hate speech when such hateful expressions do provoke racist statements (in this case, the journalist's intentions). In fact, critical race scholars stress the importance of emphasizing the impact and harm over the speaker's potentially non-existent intention to discriminate.[113] Regarding the element of the speaker's status, the Court has deemed it particularly relevant as it has ruled that politicians,[114] teachers,[115]

105 Michel Rosenfeld, *Hate Speech in Constitutional Jurisprudence: A Comparative Analysis Conference: The Inaugural Conference of the Floersheimer Center for Constitutional Democracy: Fundamentalisms, Equalities, and the Challenge to Tolerance in a Post-9/11 Environment*, 24 CARDOZO L. REV. 1523, 1565 (2002).

106 *See, e.g.*, Leroy v. France, App. No. 36109/03, ¶ 27, 31, 43 (Oct. 2, 2008), https://hudoc.echr.coe.int/eng-press?i=003-2501837-2699727.

107 *See, e.g.*, *id.*; Ceylan v. Turkey [GC], App. No. 23556/94 (July 8, 19-99), https://hudoc.echr.coe.int/fre?i=002-6560; Beizaras & Levickas v Lit huania, App. No. 41288/15 (Jan. 14, 2020), http://hudoc.echr.coe.int/eng? i=001-200344.

108 *See, e.g.*, Jersild v. Denmark, App. No. 15890/89 (July 8, 1993).

109 *See, e.g.*, Incal v. Turkey, App. No. 22678/93 (June 9, 1998), https://hudoc.echr.coe.int/fre?i=001-58197 (noting politicians enjoy a protected status, but concomitantly have heightened responsibilities in that they should avoid disseminating comments in their public speeches which are likely to foster intolerance); Féret v. Belgium, App. No. 15615/07 (July 16, 2009), https://hudoc.echr.coe.int/eng-press?i=003-2800730-3069797 (noting that politicians have the duty to refrain from using or advocating for racial discrimination).

110 *See, e.g.*, Goucha v. Portugal, App. No. 70434/12 (Mar. 22, 2016), https://hudoc.echr.coe.int/fre?i=001-161527; Feldek v. Slovakia, App. No. 29032/95 (October 12, 2001), https://hudoc.echr.coe.int/fre?i=001-59588; Ottan v. France, App. No. 41841/12 (July 19, 2018), https://hudoc. echr.coe.int/fre?i=001-182627.

111 *See, e.g.*, Gündüz v. Turkey, App. No. 35071/97 (Dec. 4, 2003), https://hudoc.echr.coe.int/fre?i=001-61522 (stating that live TV as not easy to reformulate or retract).

112 *See, e.g.*, Vejdeland & Others v. Sweden, App. No. 1813/07 (May 9, 2012), https://hudoc.echr.coe.int/eng?i=001-109046; Vereinigung Bildender Künstler v. Austria, App. No. 68354/01 (April 25, 2007), https:// hudoc.echr.coe.int/fre?i=001-79213.

113 "Good intentions are not enough," as mentioned in Henrika McCoy, *Black Lives Matter, and Yes, You Are Racist: The Parallelism of the Twentieth and Twenty-First Centuries*, 37 CHILD ADOLESC. SOC. WORK J. 463, 464 (2020), https://www.degruyter.com/document/doi/10.7208/97802 26703725/html (last visited Oct 4, 2022) (citing ANNE WARFIELD RAWLS & WAVERLY DUCK, TACIT RACISM 90 (2020)).

114 Féret v. Belgium, App. No. 15615/07 (July 16, 2009). However, in Le Pen v. France the Court did not refer to the political status of the applicant.

115 *See, e.g.*, Lilliendahl v. Iceland, App. No. 29297/18 (June 12, 2018), available at <https://hudoc.echr.coe.int/fre?i=001-203199> accessed at 21 November 2024.

and famous sporters[116] have a higher responsibility not to engage in hate speech statements. As to the reach of hateful expressions, the ECtHR has not yet directly mentioned the increased reach of online hate speech as shown by recent studies.[117]

In general terms, the ECtHR's case law on expressions that should not be protected under Article 10 because it would constitute hate speech include:

1 content that encourages violence, armed resistance or insurrection if there is (i) an intentional and direct use of wording to incite to violence and (ii) where there is a real possibility that the violence occurs;[118]
2 "glorification of terrorism" when provoking public reaction (i.e., adherence from general public to the idea), which would be capable of stirring up violence and of having a demonstrable impact on public order;[119]
3 content that creates a "pressing social need"[120] to provide protection against hateful content and, to this, Judge Tulkens suggested that the ECtHR built its argument considering that the harmful effect depended on historic, demographic and cultural contexts;[121]
4 content of racist and xenophobic nature during electoral campaigns and pronounced by members of parliament;[122]
5 serious and prejudicial allegations, even if not a direct call to hateful acts.[123]

However, the interpretation of the pressing social need in conjunction with the margin of appreciation doctrine has been controversially applied in

116 *See, e.g.*, Šimunić v. Croatia, App. No. 20373/17 (Mar. 9, 2017), https://hudoc.echr.coe.int/fre?i=001-189769.
117 Mathew et al. (n 3).
118 *See, e.g.*, Erbakan v. Turkey, App. No. 59405/00 (July 6, 2006) (finding there had been a violation of Article 10 because there was no proof of actual risk or imminent danger of the speech fomenting intolerance); Vajnai v. Hungary, App. No. 6061/10 (Sept. 23, 2014) (finding the Government failed to adduce any evidence to suggest that there is real and present danger of any political movement or party restoring the communist dictatorship); Kiraly & Domotor v. Hungary, App. No. 10851/13 (April 17, 2017) (finding authorities failed to act against racial violence and they had breached the right to respect for private life under Article 8 ECHR); Beizaras & Levickas v. Lithuania, App. No. 41288/15, ¶ 128 (Jan. 14, 2020), http://hudoc.echr.coe.int/eng?i=001-200344 (finding authorities failed to prosecute as the comments on Facebook had constituted "undisguised calls on attacks on the applicants' physical and mental integrity, which required protection by criminal law"). *E.g.*, Dicle v. Turkey, App. No. 46733/99, ¶ 33 (July 11, 2006), https://hudoc.echr.coe.int/eng?i=001-223703.
119 *See, e.g.*, Leroy v. France (n 106).
120 *See, e.g.*, I.A. v. Turkey, App. No. 42571/98 (Sep. 13, 2005), https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=001-70 113&filename=001-70113.pdf&TID=hcoelbxhnm.
121 *See, e.g.*, Soulas & Others v. France, App. No. 15948/03, ¶¶ 32-35 (Oct. 7, 2008), http://hudoc.echr.coe.int/eng?i=001-87370.
122 *See, e.g.*, Féret v. Belgium, App. No. 15615/07 (July 16, 2009), http://hudoc.echr.coe.int/eng?i=001-93626.
123 *See, e.g.*, Vejdeland & Others v. Sweden, App. No. 1813/07 (Feb. 9, 2012), https://hudoc.echr.coe.int/eng?i=001-109046.

Perinçek v. Switzerland,[124] when the Court declared a violation of Article 10 defending that in this particular context, hate speech had a diminished impact. The ECtHR looked into geographical, historical and temporal elements in the contextualization of hate speech to defend public statements in Switzerland calling the Armenian genocide a "lie" should not have been criminalized. In this judgment, as mentioned by Bayer and Bard, the Court departs from its case law in various aspects.[125]

First, the negation of genocide war crimes is to be covered as serious hate speech under Article 17 ECHR. The Court had consistently interpreted the cases of denial of the Holocaust genocide as inadmissible under Article 17 of the ECHR. Nevertheless, here the ECtHR found that the criminal conviction of the Armenian genocide denial statements was neither a "pressing social need" nor "necessary in a democratic society." This raises the question as to whether the Court would apply different thresholds for different genocides. Second, the Court fails to recognize the potential international reach of such revisionist statements. Third, this Chapter goes beyond the analysis by Bayer and Bard by also claiming this case deviates from the ECtHR jurisprudence that public figures have a special duty to not express hateful speech. *In casu*, the statements were made by a famous political figure which should have resulted in the application of the special duty to refrain from expressing hateful opinions. As a result, this case cannot but be analysed as a deviation from the European human rights standards set by the Court itself in its previous case law. Importantly, the politicians' duty to counter hate speech, including their duty to remove hateful contents posted by third-parties on the politician's social media accounts, was confirmed in a more recent ECtHR case.[126]

2.3.3 Other Treaties

Besides the ECHR, the CoE adopted other treaties that complement the regulation of hate speech. The following paragraphs present a selection of instruments impactful for the regulation of hate speech because they contain provisions on the right to non-discrimination and/or prohibitions of incitement to hatred. These treaties are introduced according to their descending order of relevance for the regulation of hate speech.

124 Perinçek v. Switzerland, App. No. 27510/08 (Oct. 15, 2015), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22002-10930%22]}.

125 Judit Bayer & Petra Bard, Hate Speech And Hate Crime In The Eu And The Evaluation Of Online Content Regulation Approaches 38 (July 2020), Https://Www.Europarl.Europa.Eu/Regdata/Etudes/Stud/2020/ 655135/Ipolstu(2020)655135_EN.pdf.

126 Sanchez v. France, App. No. 45581/15 (ECHR, 15 May 2023), https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-224928%22]}

The 2003 Additional Protocol to the Convention on Cybercrime (APCC)[127] was developed with the goal of harmonizing the criminalization of racist and xenophobic acts committed through computer systems. Such acts encompass the dissemination of racist or xenophobic material (Art. 3), racist and xenophobic motivated threats (Art. 4), insults (Art. 5), the denial, gross minimization, approval or justification of genocide or crimes against humanity (Art. 6). Aiding or abetting any of these conducts is also criminalized under this Protocol (Art 7). The APCC requires States to adopt and enforce legislation and/or other effective measures to make several types of racist conduct committed via computer systems *criminal* offences under domestic law "when committed intentionally and without right."

While representing a significant accomplishment in the protection against racism and xenophobia online, this Protocol raises further important legal debates. First, the inclusion of racist and xenophobic insults in Article 5 as a criminal act may be interpreted as to create a new standard when compared to the ECtHR jurisprudence on freedom of expression. This is because, although the ECtHR ruled expressions that "offend, shock or disturb" to be generally protected,[128] based on Article 5 of the Additional Protocol and in the event that "offensive expressions" and "insult" are interpreted as interchangeable terms, it can be argued that racist or xenophobic offenses spread through the internet can be criminally actionable. A rationale supporting this standard might be that expressions shared digitally typically reach a wider audience and this may have a far more impactful reach than when shared offline. This interpretation requires further discussion about the human rights safeguards protecting freedom of expression. Still, the acknowledgement of this heightened legal standard is a significant development in European human rights standards to counter racist and xenophobic motivated online hate speech.

Second, the APCC is not aligned with the European and international human rights standard that only the most serious cases of hate speech should be criminalized.[129] The focus on criminal law measures against online hate speech undermines the relevance of civil law or other non-legal responses as key strategies to counter and prevent further hate speech. Though the APCC provides for the possibility of not attaching criminal liability when other remedies are available and in case the conduct is not associated with hatred or violence (Art. 3(2)), this is only possible for acts of dissemination of online hate speech with racist or xenophobic intent. Furthermore, it seems unrealistic

127 Additional Protocol to the Convention on Cybercrime, Jan. 28, 2003, E.T.S. 189, https://rm.coe.int/168008160f.

128 Handyside v. UK (n 102) , ¶ 49.

129 *See, e.g.*, Eur. Ct. H.R. Press Release on Admissibility Decision of Le Pen v. France (App. No. 18788/09), and the recommendation of the Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression A/74/486 (May 7, 2010), https://hu doc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=003-3117124-34 55760&filename=003-3117124-3455760.pdf.

to apply criminal law to most cases given the high prevalence of hate speech online and the well-known length of any legal action, even more so when it is criminal. Finally, this framework heavily impacted by the aforementioned challenges associated with the almost exclusive reliance on criminal law could be said to be in violation of Article 13 ECHR, for not contributing to an effective remedy for people targeted by hate speech. A revision of this standard seems to be necessary for full compliance with European human rights law.

Third, the restriction of this Protocol to hate speech with racist or xenophobic intent leaves out other hate speech such as misogynist and queerphobic speech. The restriction to racist or xenophobic content does not align with critical legal theory nor the theory of intersectionality as it dismisses many other systems of oppression. To conclude, despite representing a significant accomplishment in elevating the threshold of protection against racist and xenophobic insults online, the APCC contributes to an overuse of criminal law, to entropy in the remedial processes for victims, and leaves unprotected many groups targeted by online hate speech.

The 2011 Convention on preventing and combating violence against women and domestic violence (Istanbul Convention)[130] also contributes to the regulatory framework of (online) hate speech. The Istanbul Convention is key in guaranteeing protection from expressions manifested offline or online with the intent to threaten the target (Arts. 33 and 34) and from unwanted verbal or non-verbal expressions of a sexual nature manifested with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment (Art. 40). Aiding and abetting stalking is also considered as an offense (Art. 41).

Three remarks are relevant with respect to this instrument's contribution to the regulation of hate speech. First, even though Article 1 of the Convention initially refers to women and victims of domestic violence, the groups protected by the selected articles are to be broadly interpreted. Not only do these provisions refer to protecting people in general but the Explanatory report[131] clarifies the drafters' wish to suggest an open-ended list of grounds for non-discrimination.[132] Suggested grounds for the open-ended list of protected

130 Convention on Preventing and Combating Violence Against Women and Domestic Violence, May 11, 2011, E.T.S. 210, https://rm. coe.int/1680 08482e.

131 Though not binding, explanatory reports are important sources for the interpretation of international law instruments.

132 Explanatory Report, Convention on Preventing and Combating Violence Against Women and Domestic Violence, ¶ 53, May 11, 2011, E.T.S. 210, https://rm.coe.int/1680a48903. See also *id*. at ¶ 87 ("For the purpose of this Convention, persons made vulnerable by particular circumstances include: pregnant women and women with young children, persons with disabilities, including those with mental or cognitive impairments, persons living in rural or remote areas, substance abusers, prostitutes, persons of national or ethnic minority background, migrants – including undocumented migrants and refugees, gay men, lesbian women, bi-sexual and transgender persons as well as HIV-positive persons, homeless

categories include among others gender, sexual orientation, gender identity, age, state of health, disability, marital status, and migrant or refugee status. Gender is also broadly conceptualized. While the main text seems to have a narrow definition in Article 3 referring to male or female only, in the Explanatory report the drafters clarified they seek to protect people on the basis of gender in a more expansive meaning when they recognize that queer people may too be persecuted on the basis of their gender.[133] The conceptualization of victims of domestic violence is also broad and applies to *all* victims (Art. 2(2)).

Second, similarly to the APCC, the Istanbul Convention prescribes that expressions creating an offensive environment should not be protected (Art. 40). Again, this could arguably be read in contradiction to the ECtHR ruling that expressions that "offend, shock or disturb" are to be generally protected. The protected groups covered from hate speech under this provision should also be interpreted as an open-ended list. Nevertheless, in line with critical legal theorists, for the expression to amount to hate speech, a contextual analysis of potential historical oppressions would need to be conducted.

Third, also in the same way as the APCC, the Istanbul Convention seems to emphasize the preference for the use of criminal law for Articles 33, 34, and 40. It should be noted that other legal civil or administrative responses could be accepted as long as these provide other means of effective, proportionate, and dissuasive measures. Still, the emphasis on criminalization should always be expressly reserved to the most serious cases, and this caveat should have been better clarified in the Istanbul Convention.

Another treaty relevant to the regulation of hate speech is the 1994 Framework Convention for the Protection of National Minorities (FCNM).[134] The FCNM prohibits discrimination and calls for equal rights before the law,[135] and it encourages intercultural dialogue as well as measures to protect persons who are subject to threats or acts of discrimination, hostility, or violence due to their ethnic, cultural, linguistic, or religious identity.[136] This Framework Convention emphasises legal action to fight bias and the need to prevent hate speech, namely through empowering the role of the media.[137]

---

persons, children and the elderly"). This broad conceptualization is also supported in Art. 12 on the general preventive measures.

133 *Id.* ¶ 53.

134 Council of Europe, Framework Convention for the Protection of National Minorities and Explanatory Report, European Treaty Series – No. 157, Doc. H9510 (1995), available at <https://rm.coe.int/168007cdac> accessed 22 November 2024.

135 FCNM (n 134), Art. 4.

136 FCNM (n 134), Art. 6.

137 FCNM (n 134), Art. 9.

2.3.4 Non-Treaty Initiatives

There are also non-treaty instruments at the level of the Council of Europe, which contribute to defining the legal contours of hate speech in Europe. These instruments were selected on the basis of containing either direct references to hate speech or references to discrimination, tolerance, or to the protection of marginalized groups. Some instruments specifically regulate online harms, including online hate speech.

The study of the responsibilities of private actors when regulating online hate speech (procedural regulation) is not the main focus of this Chapter and will, therefore, not be developed in detail. Though some initiatives on business and human rights will be mentioned for ease of reference on instruments impacting the regulation of online hate speech, this Chapter focuses instead on clarifying the main elements in the conceptualization of hate speech (substantive regulation).

The non-treaty framework at the CoE is presented following the legal status of the sources and the subsequent order of influence of the instruments for the case-law of the ECtHR. Firstly, this Section will focus on the Recommendations and Guidelines of the Committee of Ministers (CM), and secondly, General Policy Recommendations of the European Commission against Racism and Intolerance (ECRI). The instruments produced by these bodies are the most important non-treaty initiatives because they are frequently mentioned by the ECtHR in its case law, and, as noted by McGonagle, the Court recognizes these bodies produce standard-setting work.[138] Subsequently, this Section will describe the most relevant communications at the CoE level countering hate speech.

The Committee of Ministers (CM) adopted three recommendations directly guiding the regulation of hate speech, two recommendations on measures to combat discrimination, and three recommendations with guidelines on how to regulate business and human rights, the latter particularly relevant for the regulation of online hate speech.

Starting with the recommendations directly guiding the regulation of hate speech, Recommendation No. R (97) 20 was the first direct and expansive communication from the CM on the topic of hate speech. This Recommendation defines hate speech as "all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism 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" and notes that governmental representatives must both refrain from hateful statements and establish a legal

138 As noted by McGonagle, the ECtHR recognized that both the CM and the ECRI manage to engage with specific topics in a more expansive manner while taking into consideration current practice and potential developments. *See* McGonagle (n 17), 27.

framework for civil, administrative and criminal responses.[139] The principles to combat hate speech stated in this recommendation place a significant emphasis on the need to have a legal framework where multiple stakeholders are responsible for contributing to the monitoring and countering of hate speech, highlighting the special responsibility of public officials and the media.[140]

Recommendation No. R (97) 21 was adopted immediately after Recommendation No. R (97) 20 and focuses on the need to capitalize on contributions by the media to prevent hate speech by promoting a culture of tolerance. This recommendation highlighted that, in order to counter hate speech, it was not enough to respond with legal measures and underscored the importance that the media adopted programs to promote access to media for minorities and codes of conduct promoting tolerance. An additional emphasis was placed on the importance to train media professionals on multiculturalism and to promote integration and airtime for all individuals, especially ensuring access and representation of marginalized communities.

On 20 May 2022, the CM adopted Recommendation CM/Rec(2022)16 in a wide-ranging strategy to combat hate speech in light of current challenges brought about by technological developments and the rise of hate speech prevalence, especially on social media. This recommendation was prepared by a Committee of Experts (ADI/MSI-DIS) that had been set up in 2020 by the CM with the goal of drafting an updated framework for a comprehensive human rights strategy to address hate speech, including in the online environment. The main findings and suggestions in CM/Rec(2022)16 build on existing CoE treaties and other standard-setting initiatives as well as on the ECtHR case law. This recommendation is a landmark instrument at the CoE level as it has the potential to provide a clear and updated roadmap for the regulation of hate speech in the broader European context. A dedicated reflection is included below in Section 2.3.4.1.

Two other CM recommendations address hate speech in the form of discrimination. Recommendation (2010)5 on measures to combat discrimination on grounds of sexual orientation or gender identity, expressly included the obligation to combat inciting hatred or other forms of discrimination against

139 Recommendation of the Committee of Ministers on Hate Speech, Doc. R 97 (1997).

140 Principle 1 specifies that public officials are under a special responsibility to refrain from stating or inciting hate, particularly in the media. Principles 2, 3 and 4 reinforce the idea that States must guarantee a legal framework composed of civil, administrative and criminal law to address hate speech, with the caveat that only the most serious hateful expressions should be criminalized. Additionally, in Principle 6, the CM differentiated between the responsibility of the author of hate speech and the responsibility of the media reporting such hate speech, underlining that the latter must be able to communicate on matters of public interest such as the abuse of freedom of expression through hate speech. Regarding Principle 6, this research argues though that this principle should not apply to online platforms with algorithm designed to promote hate speech given its virality. For further info, *see* McGonagle (n 17), 23.

the LGBTI+ community. In Recommendation (2011)7 on a new notion of media, the CM stated that digital platforms should monitor the use of biased expressions and defended that these actors be required by law to report, to the competent authorities, criminal threats of violence based on racial, ethnic, religious, gender or other grounds that come to their attention.

There are also three other CM recommendations on business and human rights with particular relevance for the regulation of online hate speech. Recommendation (2016)3 on human rights and business; Recommendation (2018)2 on the roles and responsibilities of internet intermediaries; and Recommendation (2020)1[141] on the impacts of algorithmic systems on human rights, together lay the groundwork for corporate responsibility to comply with human rights safeguards needed to prevent online hate speech through digital means.[142]

The CM has also in recent years produced two other instruments impacting the regulation of hate speech, which even if not in the form of recommendations, still reflect the views of the Committee and can thus influence the case-law of the ECtHR. In 2019, the CM adopted a Declaration on the manipulative capabilities of algorithmic processes[143] warning against the risk of using algorithmic processes to manipulate social and political behaviour. And in 2021, the CM adopted the Guidelines on upholding equality and protecting against discrimination and hate during times of crisis.[144] These guidelines alert the need to counter online hate speech during times of crises, namely by improving data collection on marginalized groups; by involving the affected community in this research; by providing continued access to information, legal, psychological and social support to victims; and by having public authorities as role models publicly rejecting hate speech. In a remarkable consideration, these guidelines also recognize the intersectional discrimination of hate speech.[145]

---

141 Committee of Ministers, Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems, Council of Europe, https://search.coe.int/cm/pages/result _details.aspx?objectid=09000016809e1154.

142 *Id.* Overall, the CM underscores the need to support national institutions in the oversight, risk assessment and enforcement of the United Nations Guiding Principles on Business and Human Rights by business running algorithmic systems.

143 Declaration by the Comm. of Ministers on the Manipulative Capabilities of Algorithmic Processes (Feb. 13, 2019), https://search. coe.int/cm/pages/result_details.aspx?ObjectId=090000168.

144 Steering Comm. on Anti-Discrimination, Diversity And Inclusion (CDADI), Guidelines of the Committee of Ministers of the Council of Europe on Upholding Equality and Protecting Against Discrimination and Hat e During the Covid-19 Pandemic and Similar Crises in the Future, Council of Europe (May 2020), https://policycommons.net/artifacts/1811064/guidelines-of-the-committee-of-ministers-of-the-council-of-europe-on-upholding-equality-and-protecting-against-discrimination-and-hate-during-the-covid-19-pandemic-and-similar-crises-in-the-future-2021/2547081/ on 05 Mar 2023. CID: 20.500.12592/0sjj6g.

145 CDADI (n 144).

The European Commission against Racism and Intolerance (ECRI) has also engaged specifically with the topic of hate speech. The most prominent General Policy Recommendation (GPR) is GPR No. 15 on combating hate speech, where ECRI defines hate speech as "advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of 'race,' colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status." The use of the word "insult" for the typification of the hate speech acts, despite having to be firstly interpreted in light of national legislative provisions, could also raise questions of potential lack of coherence with the ECtHR Handyside judgment.[146] However, it is important to note that both the APCC and the Istanbul Convention already utilized respectively, racial and xenophobic insults and sexual harassment conducive to offensive environment as some of the most serious forms of hate speech, even suggesting their criminalization. Thus, one possible interpretation here is that insults addressed at historically oppressed groups on the basis of their race or gender (non-exclusive interpretation), would be considered a seriously grave form of hate speech. Notably, ECRI GPR 15 promotes self-regulation of media and states that criminalization is necessary following the Rabat Plan of Action,[147] i.e., in circumstances where hate speech is intended or can reasonably be expected to incite acts of violence, intimidation, hostility or discrimination against those targeted.

ECRI adopted three other GPRs relevant for this analysis. GPR 7 on national legislation to combat racism and racial discrimination contributed to remarkable developments for the regulation of hate speech when it declared that intent to incite the commission of acts of violence, intimidation, hostility or discrimination is not essential to criminalize. Instead, it indicated that criminal law can also be used when violence, intimidation, hostility or discrimination can reasonably be expected to be the effect of using the hate speech concerned and therefore the use of such speech would be considered reckless. GPR 6 on combating the dissemination of racist, xenophobic and antisemitic material via the Internet requested governments to take the necessary measures at national and international levels to act effectively against the use of the internet for racism, xenophobia, and antisemitism. GPR 11 on combating racism and racial discrimination in policing, ECRI asks States to ensure law enforcement investigates racist offenses in victim-friendly environments.

---

146 *See* Handyside v. UK (n 102). In the Handyside judgment the Court posited that freedom of expression applies also to ideas that offend, shock or disturb the State or any sector of the population.

147 Annual Report. of the U.N. High Commissioner for Human Rights: addendum, ¶ 34, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013).

Also relevant for the analysis of non-treaty initiatives on hate speech at the CoE is the 2008 Report by the European Commission for Democracy through Law (Venice Commission) on the relationship between freedom of expression and freedom of religion.[148] In this report, the Venice Commission defends incitement to hatred, including religious hatred, should have a specific requirement of the intention of recklessness to be criminalized and concludes the offense of blasphemy should be abolished.

Finally, the European Ministerial Conferences on Mass Media Policy and Council of Europe Conferences of Ministers responsible for Media and New Communication Services reiterate in various instances the need to tackle hate speech and promote tolerance. Of note, in 2013, the Ministers responsible for Media and Information Society decided "to protect people from the risks encountered on the Internet, in particular by fighting cybercrime, sexual abuse, exploitation of children, cyberbullying, gender-based discrimination, incitement to violence, hatred and any form of hate speech," and invited the CoE to "continue to combat hate speech and incitement to violence and terrorism, whether involving individuals, public or political persons or groups, including offering guidance on ways to mitigate its escalation, due to the speed and scope of its online dissemination."[149]

2.3.4.1 Observations Regarding the CM Recommendation on Combating Hate Speech

This Section provides a dedicated analysis of CM/Rec(2022)16[150] as this recommendation will hopefully pave the way for a comprehensive and renewed effort on the regulation of hate speech in the European context. For instance, this recommendation could influence the European Union's developing position on the regulation of hate speech – the European Commission communicated in December 2021 its intent to extend the list of the EU crimes to hate speech and hate crime.[151] This analysis will be divided into three main segments: (1) considerations related to the people targeted by hate speech, (2) considerations on the conceptualization of hate speech and (3) its regulation, and con-

148 Council of Europe, Report by the European Commission for Democracy through Law (Venice Commission) on the relationship between freedom of expression and freedom of religion, CDL-AD(2008)026 (2008).

149 Council of Europe, European Ministerial Conferences on Mass Media Policy and Council of Europe Conferences of Ministers responsible for Media and New Communications Services, https://rm.coe.int/168064 61f b (2013).

150 Council of Europe, Recommendation of the Committee of Ministers to Member States on Combating Hate Speech, CM/Rec(2022)16, https://se arch.coe.int/cm/Pages/result_ details.aspx?ObjectId=0900001680a67955.

151 European Parliament, Proposal to Extend the List of EU Crimes to All Hate Crimes and Hate Speech, COM(2021)0777, https://www.europa rl.europa.eu/legislative-train/theme-a-new-push-for-european-democracy/f ile-hate-crimes-and-hate-speech#:~:text=On%209% 20December%202021 %2C%20the,out%20in%20Article%2083(1) (2021).

siderations on the online applicability of the legal framework regulating hate speech.

First, regarding the representation of people targeted by hate speech, it is positive to note the Recommendation clarifies the legal framework for support of those targeted by hate speech. The first accomplishment is that CM/Rec(2022)16 refrains from employing the expression "victims of hate speech". This is a valuable development as there has been a growing advocacy work done in the past decade alerting the use of the word "victim" suggests an inescapable position of fragility, which may result in double victimization of those targeted by hate speech. It is therefore recommended to use the expression "people targeted by hate speech."[152] The second accomplishment is that this Recommendation shares an official disclaimer that, even if race is included as a potential factor in discriminatory practices,[153] this is a highly contested concept. Still, the recommendation could have gone beyond this position by acknowledging race is a concept created during the colonial period by white supremacist colonial powers as a means to justify their colonial project characterized by discriminatory and criminal policies.

CM/Rec(2022)16 also remarkably reflects ongoing debates about the prevention of marginalization practices of minority groups when it requires that MS ensure translations into minority languages,[154] about the cumulative effect of systematic and long-term exposure to hate speech[155] and about the increased harm of hate speech when considering the intersectionality of different systems of oppression for people targeted by hate speech.[156]

Nevertheless, two points could have been addressed differently to better represent those targeted by hate speech. First, despite being understandable that referring to age and gender-sensitive hate speech policies serves illustrative purposes,[157] using "all-inclusive approaches" would have better reflected the intersectionality of groups targeted by hate speech.[158] Similarly, when referring to the queer community, it would have been more inclusive to use queerphobia instead of "LGBTI" (para. 11(d)). LGBTI does not include people who are questioning or who do not identify with other labels but are still within the

152 This research acknowledges this debate and aligns with the advocates of the expression "people targeted by hate speech." However, given the need for an extensive utilization in this research and given the fact that the EU legal system uses "victim" in its EU Victims Directive, for these practical reasons this research will use victims.

153 CM/Rec(2022)16, para. 1(2).

154 CM/Rec(2022)16, Preamble.

155 CM/Rec(2022)16, para. 1(6)(d).

156 CM/Rec(2022)16, para. 58.

157 CM/Rec(2022)16, para. 1(6)(d) and para. 58.

158 The monitoring rounds of the EU Code of Conduct on countering illegal hate speech online show that the main reported types of hate speech in Europe are sexual orientation, xenophobia and hate towards the Roma community. *See* 6th Evaluation of the Code of Conduct, (Oct. 7. 2021), https://ec.europa.eu/info/sites/default/files/factsheet-6th-monitoring-round-of-the-code-of-conduct_october2021_en_1.pdf, 4.

main targets of hate speech, and "queer" is often used as an umbrella term to refer to all LGBTI people and would have thus been a more inclusive term.[159]

Second, in regard to the conceptualization and regulation of hate speech, this Recommendation provides legal clarity in many aspects. First, although there is no specific human right protecting people from being targeted with hate speech, the Recommendation clarifies in its Preamble that the regulation of hate speech requires a careful balance between the right to private and family life,[160] the prohibition of discrimination[161] and the right to freedom of expression.[162] Notwithstanding, even if the prohibition of abuse of rights[163] is also mentioned in the Preamble, it could have also been mentioned together at the start of the Preamble to better explain the ECtHR legal framework on hate speech. It would have been clearer to explain the legal system altogether, i.e., either the hateful expression is the most serious form of hate speech and therefore an abuse of right under Article 17 ECHR, or it is a violation of the right to private and family life (Article 8 ECHR) and of the prohibition of discrimination (Article 14 ECHR) and therefore should be limited as per the conditions to limit freedom of expression contained in Article 10(2) ECHR. Finally, when referring to Article 17, the Recommendation only mentions that it covers expressions aiming at destroying any rights in the ECHR.[164] But as per the case law of the ECtHR, Article 17 ECHR also covers expressions that aim to limit the rights in the ECHR beyond the limitations allowed for in the ECHR.[165]

Additionally, this Recommendation explains the three types of harmful expressions and clarifies the relation with hate speech.[166] To simplify, it explains that hate speech is a term informally used across many research fields to cover a spectrum of expressions including: (a) criminal offences; (b) problematic expressions, which although not criminal offenses, could be prohibited under civil or administrative law; and (c) content which bears no legal implications but still raises issues of respect and tolerance and should be addressed through culture and education. However, and quite importantly, in legal terms, this Recommendation confirms in paragraph 1(3) that the term "hate speech" should only be used for expressions that are (a) criminalized, and for expres-

159 The New York Times (2022), Using the Word 'Queer" Instead of 'Gay', available at <https://www.nytimes.com/2022/11/13/opinion/letters/lgbt-gay-queer.html> accessed 22 November 2024; Heather Love, "Queer." *Transgender studies quarterly* 1.1-2 (2014): 172-176.

160 ECHR, Art 8.

161 ECHR, Art. 14.

162 ECHR, Art. 10.

163 ECHR, Art. 17.

164 CM/Rec(2022)16, Preamble para. 12.

165 *See also, Factsheet – Hate Speech*, Press Unit, European Court of Human Rights (June 2022).

166 This explanation aligns with the United Nations' view on the relation between hate speech and harmful speech. *See* Ann. Rep. of the U.N. High Comm'r for Hum. Rts.: addendum, ¶ 12, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013).

sions that, although not amounting to criminal offenses, are (b) prohibited if restricted in line with the conditions laid down in Article 10(2) ECHR.

This distinction is a key contribution from this instrument as it justifies why no legal text should refer to hate speech as illegal or legal. In fact, this means hate speech is always illegal and victims can always seek legal redress, thus there is no need to include "illegal." Still, when referring to applicability of criminal law in the Preamble, it would have added legal clarity had the text mentioned that only the most serious forms of hate speech should be criminalized, and not "some." It is also positive that this Recommendation recognizes the importance of having a comprehensive system composed of various types of measures to address hate speech. It does not only make reference to criminal, civil and administrative law, but it also directs to education and training on human rights, especially as a means to respond to expressions with no legal implication, which still raises issues of tolerance and respect (para. 1(3)(b)).

Additionally, this instrument clarifies the legal framework to assess the severity of the hateful expression as well as the liability framework by referring not only to the conditions for the restriction of the right to freedom of expression as per Article 10(2) ECHR but also to the contextual variables[167] evaluated by the ECtHR on hate speech cases (para. 1(4)). Still, it would have been important to acknowledge the following elements in the description of the contextual variables: the need to investigate historical oppressions as part of the political and social contexts; the need to expressly consider the intersectionality impact of hate speech; and finally, the debate about the element of intent and how some scholars question its relevance for context analysis.[168] Regarding the latter, the element of intentionality was, however, challenged in the dissenting opinion by Judges Ryssdal, Bernhardt, Spielman and Loizou in Jersild v. Denmark when they admitted that good intentions are not enough when provoking racist statements (related to a journalists' intention in the case). This is because, oftentimes, intention is an element that is difficult to assess and also because the harm through the incitement to violence, discrimination or hatred lies precisely in the incitement against democratic and human rights values.

This Recommendation also clarifies the legal threshold for an insult to be considered criminalized hate speech when in paragraph 11(d) it clarifies that racist, xenophobic, sexist and LGBTI-phobic insults are amongst the most serious forms of hate speech and therefore subject to criminal liability. This is a very relevant explanation because it changes the usual doctrine narrative about the conditions for the restriction of freedom under the ECHR. While

167 *See* Section Part 3.2.3.

168 Katharine Gelber (2021) Critical review of International Social and Political Philosophy, 24:4, 402, available at <https://doi.org/10.1080/13698230.2019.1576006> accessed 22 November 2024.

any legal debate about freedom of expression has since the Handyside judgment started by highlighting that freedom of expression encompasses ideas that "offend, shock or disturb,'' it is now clear that in the case of racist, xenophobic sexist and LGBTI-phobic offenses or insults these are to be considered hate speech under criminal law. This threshold follows the legal framework provided for in the APCC and in the Istanbul Convention, thus increasing the legal coherence at the Council of Europe amongst instruments regulating hate speech.

Another achievement in this recommendation is the inclusion of a clear identification of the various stakeholders involved in the regulation of online hate speech, i.e., internet intermediaries, public officials and bodies, media professionals, and civil society organizations. Still, the responsibilities of some stakeholders are vague or described in a way that does not fully reflect the ECtHR case law. For instance, in reference to the responsibilities of public officials, the Recommendation merely suggests that they *avoid* engaging in hate speech (paras. 28 and 29) while the Court's position is that public figures, and especially politicians, have the *duty to refrain* from engagement in hate speech.[169] A good development is that the Recommendation expressly acknowledges that the positive obligations of Member States to prevent human rights violations also apply in the digital environment (Preamble).

Third, regarding the main elements of this Recommendation with implications for the regulation of hate speech in the online environment, it is positive to note that, when discussing the contextual variables to assess the severity of the hateful expression, one of the elements to consider is how the expression is disseminated or amplified. This clear recognition is essential to adequately represent the usually increased damage caused by online hate speech given its elevated reach and fast dissemination.

Turning to content moderation practices, the main consideration is the different threshold between paragraph 16, where the requirement is that Member States remove all hate speech offline or online, and paragraph 32, requiring internet intermediaries to remove only the most serious cases of online hate speech. This can perhaps be understood based on the premise that Member States are obliged to comply with human rights, while the human rights framework is a priori not legally binding upon internet intermediaries. Following this rationale, it is a good legal strategy to require online platforms to remove only the most serious cases of hate speech to avoid over-removal of content, which may not be legally considered hate speech. Nevertheless, it remains unclear how internet intermediaries should moderate the less severe cases of hate speech. In this regard, this Chapter suggests that the CM/Rec(2022)16 could have provided better guidance for the less severe cases by suggesting, for example, that such hate speech should be blocked, labelled

169 *See, e.g.*, Féret v. Belgium, App. No. 15615/07, 573 (July 16, 2009).

and communicated to the respective public oversight body or law enforcement for investigation. To clarify, this Chapter concurs that platforms should remove hate speech actionable under criminal law and suggests platforms should block and label hate speech actionable under civil or administrative law. The CM/Rec(2022)16 could have set this framework more expressly at least for the larger internet intermediaries.

The Recommendation also provides more general guidance on business models and content moderation. It is positive to note the suggestion of the decentralization of content moderation practices because this is a means to better achieve a good contextualization of the expression (para. 33). Additionally, it is noteworthy that even if automation or artificial intelligence systems are deployed, these should still be overseen by human moderation (para. 33) as well as the requirement that human moderators receive training to be up to date with human rights standards (para. 34). Finally, it is remarkable that the Recommendation requires internet intermediaries to *ensure* their business models are not grounded on strategies that directly or indirectly increase hate speech prevalence (para. 36). Including a requirement for businesses to proactively ensure algorithms do not promote hate speech is a completely innovative and needed[170] legal strategy.

With respect to the cooperation between internet intermediaries with law enforcement, the Recommendation in paragraph 22 seems to require that internet intermediaries report only criminal hate speech to law enforcement. Still, if this provision is read in conjunction with paragraph 19 where it is prescribed that all online hate speech be reported to public authorities, it should be concluded that internet intermediaries should report any hate speech to the competent public authorities. This cooperation is all the more important given the requirement in paragraph 16 that Member States and public authorities should remove all hate speech offline and online. Moreover, it is also important to have effective cooperation between internet intermediaries and public bodies to ensure Member States report on hate speech statistics as required in paragraph 25. Notwithstanding, the Recommendation could have clarified the need to have standardized indicators to study hate speech statistics. In fact, the disconnect between the disaggregated data is one of the main current challenges with the reporting as part of the monitoring rounds for the European Code of Conduct on countering illegal hate speech online. This would have presented a good opportunity to provide guidance to European society.

There are also special considerations concerning the responsibilities of internet intermediaries to support those targeted by hate speech. It is positive to note the Recommendation focuses on remedial processes ultimately facili-

170 *See, e.g.*, Jim Waterson & Dan Milmo, *Facebook Whistleblower Frances Haugen Calls for Urgent External Regulation*, Guardian (Oct. 25, 2021),-https://www.theguardian.com/technology/2021/oct/25/facebook-whistleblower-frances-haugen-calls-for-urgent-external-regulation.

tated through independent judicial reviews (para. 20) and that it requires short and concise explanations to all affected by online hate speech moderation practices (para. 23). However, the remedial responsibilities of internet intermediaries could have been emphasized under section 5 of the Recommendation. For example, involving internet intermediaries in facilitating counter-narratives, or even in cases where the online platform facilitated the dissemination of hate speech they are themselves to be held liable, to provide adequate reparations to those targeted.

In terms of the liability regime for the internet intermediaries, the Recommendation misses the opportunity to address the current preference of some Member States for the regulation of online hate speech through the imposition of fines upon internet intermediaries. This practice can lead to limiting the right to freedom of expression beyond the limits prescribed by ECHR and can result in limiting access to the information and to the media or in worst cases limit the protection of human rights activists.[171] It is positive to see a reference to the importance of considering the various sizes and types of internet intermediaries (para. 21).[172] However, the impact of the different sizes of internet intermediaries for the liability framework could have been more clearly explained. For example, according to the type of platform, different requirements in terms of content moderation practices could be adopted.

2.3.5 Overview Council of Europe, Hate Speech and Historical Oppression

In summary, when analysing the jurisprudence of the Court on Article 10(2) in the light of the original conceptualization by Matsuda[173] of hate speech as expressions perpetuating historical or systematic oppressions, two remarks are in order. First, though it should be recognized that the Court does include political and social background considerations as contextual variables, it fails to consistently refer to historical or systematic oppressions as a key element of hate speech. Similarly, though the Court acknowledges the importance of taking into account the victims' perspectives, it fails to consistently recognize the intersectionality of systems of oppression perpetuated by hate speech.

Regarding other treaties regulating hate speech, these broadly align with the jurisprudence of the ECtHR on the regulation of hate speech and, in some cases, even go beyond to expressly acknowledge the list of protected categories

171 Avi Asher-Schapiro & Ban Barkawi, *'Lost Memories': War crimes evidence threatened by AI moderation*, REUTERS (June 19, 2020), https://www.reuters.com/article/us-global-socialmedia-rights-trfn-idUSKB N23Q2TO.

172 Note the need to take into account the different sizes and functions of internet intermediaries in the design of the liability framework also aligns with the legal framework established by the DSA in the EU.

173 The conceptualization of (racist) hate speech introduced by Matsuda is explained in Section 2 of this Chapter.

as open-ended (e.g. the Istanbul Convention). This recognition of the open-ended conceptualization of the protected groups aligns with the intersectionality theory and with critical legal theory. Still, these treaties fail to expressly mention the element of historical oppression key to the identification of groups targeted by hate speech.

In reviewing the non-treaty initiatives in light of the original conceptualization of hate speech, these initiatives fail to *formally* establish the historical and intersectional elements of oppression in hate speech conceptualized by critical race scholars as part of the contextual variables to be assessed in hate speech cases. Still, progress towards legal coherence has been made and the progressive critical approach adopted in two of the initiatives in mentioning the intersectionality of different systems of oppression, i.e., CM guidelines on upholding equality and protecting against discrimination and hate during times of crisis and CM/Rec(2022)16, should be acknowledged.

2.4 Approaches to Hate Speech by the European Union

2.4.1 General Principles and Primary Sources

The European Union is built on a set of values prescribed in Articles 2 and 3 of the Treaty of the European Union (TEU), which the Member States (MS) resolve to respect and promote. Article 2 of the TEU explains that the foundational values for membership to the EU are human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of minorities. Article 3 clarifies that the EU shall offer its citizens freedom, security and justice.

Although the EU had historically focused on economic integration, it soon acknowledged the growing need to promote human rights.[174] The Court of Justice of the European Union (CJEU) noted in the 1970s that the protection of human rights was a general principle of law for the EU and confirmed that the EU took the ECHR as its main source of inspiration.[175] In Article 6(2) of the TEU, the EU commits to accede to the ECHR, to take the rights in the ECHR as its general principles, and to follow the interpretation of the ECtHR.

The alignment of the EU with the ECHR is also prescribed in the Charter of Fundamental Rights of the EU (CFREU). The CFREU is the EU's leading treaty for the protection of human rights and therefore a key instrument in the regulation of hate speech. The CFREU seeks to ensure European coherence when in Article 52(3) it contains that for articles with corresponding rights

174 Case 29/69 Stauder v. Stadt Ulm, 1969 ECR 419.

175 CJEU, Case *11-70 Internationale Handelsgesellschaft*, paragraph 4; Case *C-60/84 Cinéthèque*, paragraph 26, available at <https://e-justice.europa.eu/563/EN/part_i__protecting_fundamental_rights_within_the_european_union#p1s2.3> accessed 22 November 2024.

in the ECHR, "the meaning and scope of those rights shall be the same"[176] and, likewise, the jurisprudence of the ECtHR over such rights is also applicable to the EU. It should be noted that the EU reserves the right to grant a higher level of protection of human rights than the protection conferred by the ECHR. The complementarity between the EU and the CoE for the protection of human rights has been extensively debated and, in 2005, the CoE suggested the EU should even transpose aspects of other CoE Conventions when within its competence as per the EU Law.[177]

The EU is specifically committed to fighting discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation when designing and implementing its policies (Art. 10 of the Treaty on the Functioning of the European Union (TFEU)). This principle is reinforced in Article 19 of the TFEU, which enables the EU to "take appropriate action to combat discrimination." Additionally, Article 67(3) of the TFEU stipulates that the EU must "ensure a high level of security through measures to prevent and combat crime, racism and xenophobia, and through measures for coordination and cooperation between police and judicial authorities and other competent authorities, as well as through the mutual recognition of judgments in criminal matters." It should be noted that in order to achieve this goal, the EU may adopt measures to approximate criminal laws, namely on hate speech. In fact, the European Commission (EC) communicated its intent to extend the list of the EU crimes to hate speech and hate crime.[178]

Applying this framework to CFREU articles impacting the regulation of hate speech (e.g. Article 1 on human dignity, Article 11 on freedom of expression, and Article 21 on the right to non-discrimination), the CJEU must follow as the minimum standards the conditions for limitation of freedom of expression as per Article 10(2) and adhere to standards developed by the ECtHR in related jurisprudence.[179] For example, due to the wording "such as," the prohibition of discrimination in the CRF (Art. 21 CFR) reflects an open-ended list of impermissible grounds for discrimination, which is also mirroring the formulation in the ECHR (Art. 14). While the CFREU contains an open-ended list of impermissible grounds for discrimination, it also contains specific provisions on certain impermissible grounds such as the respect of diversity (Art. 22), gender equality (Art. 23), age (Arts. 24 and 25), and disabilities (Art. 26).

176 Tarlach McGonagle, *Chapter 24 Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation*, 17 (pre-publication).

177 Council of Europe, Ministers' Deputies CM Documents, Action Plan, CM(2005)80 final 17 May 2005, Appendix I, Guideline 5.

178 Council of Europe, CDADI (n 144).

179 EU Network of Independent Experts on Fundamental Rights, Commentary of the Charter of Fundamental Rights of the European Union (2006) 400.

2.4.2 Secondary Sources

Focusing now on the secondary sources of law of the European Union regulating hate speech, this Section highlights, first, both the Council Framework Decision on combating certain forms and expression of racism and xenophobia by means of criminal law (Framework Decision)[180] and the Audiovisual Media Services Directive (AVMSD). After that, this Section expands on two resolutions adopted by the European Parliament (EP) relevant to explain the regulatory framework on hate speech and to contextualize the ascension of hate speech in the agenda of the EU. This Section also explores three legislative avenues:[181] (1) the EC Regulation of the EP and of the Council on a Single Market for Digital Services (Digital Services Act, DSA);[182] (2) the Regulation of the EP and of the Council Laying down harmonized rules on artificial intelligence (Artificial Intelligence Act, AI Act);[183] and (3) the Directive of the EP and of the Council on combating violence against women and domestic violence. This Section will also explore the intent of the EC communication from December 2021 to extend the list of the EU crimes to hate speech and hate crime.

Similar to the analysis of non-treaty initiatives at the Council of Europe, the study of the responsibilities of private actors when regulating online hate speech (procedural regulation) will not be extensively developed. Although some initiatives on business and human rights will be mentioned for ease of reference on instruments impacting the regulation of online hate speech, this Chapter focuses instead on clarifying the main elements in the conceptualization of hate speech (substantive regulation).

The Framework Decision criminalizes two types of speech – (1) publicly inciting to violence or hatred and (2) publicly condoning, denying or grossly trivializing crimes of genocide, crimes against humanity and war crimes – when they are directed against a group of persons or a member of such a group defined by reference to "race, colour, religion, descent or national or ethnic origin." From a European human rights perspective, it is problematic that the list of protected grounds is limited to "race, colour, religion, descent or national or ethnic origin." Still, the Framework Decision explicitly suggests that Member States may adopt provisions in national law of crimes "against

180 Acts Adopted Under Title VI of the EU Treaty: Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, 328 OFF. J. OF THE EUR. UNION 55, 55–58 (2008).

181 At the time of publishing, these three legislative acts had not been adopted yet and their analysis had thus been grouped under 'legislative avenues'.

182 European Union, Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

183 *Europe: Artificial Intelligence Act Must Protect Free Speech and Privacy*, ARTICLE19, https://www.article19.org/resources/europe-artificial-intelligence-act-must-protect-freedom-of-expression-and-privacy/.

a group of persons defined by other criteria than race, colour, religion, descent or national or ethnic origin, such as social status or political convictions"[184] This wording does not, however, refer to misogynistic or queerphobic speech. In conclusion, the best regulatory strategy to ensure protection of all people targeted by hate speech would have been to leave the list officially open-ended by following Article 21 of the EU CFR. Had the EU legislator chosen an explicit open-ended clause, it would have certainly increased the legal protection of marginalized groups from hate speech. Similarly, though the Framework Decision also calls upon its Member States to guarantee that racist or xenophobic motivation is considered an aggravating factor in criminal law (Art. 4), this approach also does not explicitly mention queerphobia or misogyny motivations.

Another remark is that the Framework Decision lays the emphasis on criminalization, which may wrongfully suggest that other responsive and preventative means should not be prioritized.[185] Similar to the previous analysis on protected groups, even though this instrument does allow for a margin of consideration of severity (Article 2 specifies that Parties may choose to punish only conduct that is likely to disturb public order or which is threatening, abusive or insulting), it would have certainly been a better legislative strategy to include the possibility of various legal (including civil and administrative) and educational measures upfront. Furthermore, the prioritization of criminal law undermines the frequent scepticism of hate speech victims to report it to law enforcement entities for fear of double victimization. To improve the protection of victims within the scope of the Framework Decision, the EC called for the complementary application of the Victims Directive.[186]

The AVMSD governs EU-wide coordination of national legislation on all audiovisual media – traditional TV broadcasts, on-demand services and video-sharing platforms. As per its latest review from 2018, the VMSD recognizes in Recital 45 the problems of increased hate speech and harmful content online. Article 28b states that video-sharing platforms should be required to "take appropriate measures to protect the general public from content that contains incitement to violence or hatred directed against a group or a member of a group on any of the grounds referred to in Article 21 of the CFREU or the dissemination of which constitutes a criminal offense under Union law." As seen, Article 21 of the CFREU postulates a broad understanding of impermis-

184 Acts Adopted Under Title VI of the EU Treaty (n 180), at 56, Preamble, Para. 10.

185 Despite the reference in the Preamble, Para. 6.

186 Report from the Commission to the European Parliament and the Council on the implementation of Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law, COM (2014) 27 final, 9 (Nov. 2018).

sible grounds[187] for discrimination including "sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation."

Nevertheless, despite prescribing that video-sharing platforms take "appropriate measures" to protect the general public from hate speech on the broad accounts as per Article 21 CFR, the AVMSD only requires Member States to make criminal action available for victims of hate speech as per the Framework Decision, which as just mentioned above applies in racist and xenophobic hate speech cases only (Art. 28b(1)(I)).

The EP adopted two resolutions concerning the protection from hate speech raising attention to the importance of strengthening legal protections from hate speech, especially for certain groups. EP Resolution 2013/2543(RSP) strengthens, inter alia, the need to continue extending protection from discrimination on the basis of non-exhaustive grounds (para. F). EP Resolution 2019/2933(RSP) reinforces the need to adopt an expansive consideration of impermissible grounds for hate speech and underlines the need to address offline and online hate speech (para. 8). This Resolution also calls on Member States to develop mechanisms for monitoring, reporting and investigating online and offline hate speech.

At the EU level, there are three main instruments shaping the regulation of online hate speech.[188] First, the regulation on the DSA, which amends the e-Commerce Directive. The DSA aims to update and harmonize across the EU content moderation practices, due diligence rules, and the framework on the liability of internet intermediaries. Although it reinforces the importance to respect fundamental rights, the DSA does not define what is illegal content. This instrument is what is called a horizontal baseline regime as it focuses on procedural safeguards. For the definition of illegal content in the EU, it is essential to analyse the vertical hard-law *lex specialis*. In the case of hate speech, the *lex specialis* is for the time being the Framework Decision. Addi-

187 This phrasing is inspired by the work of Tarlach McGonagle, *Minority Rights, Freedom of Expression and of the Media: Dynamics and Dilemmas, School of Human Rights Research Series*, 44. The present work will refrain from using "protected categories" and opt instead for "impermissible grounds" because the first can lead the reader to assume that certain segments of the population are inherently disempowered and can thereby contribute to a wrongful stigmatization of such identities, which this research strongly wishes to contest. Hence, in search for a more accurate lexicon, this study chooses to utilize "impermissible grounds" to put the emphasis on the role of law to counter hatred and in an effort to lift any stigmatization of traditionally targeted groups.

188 It is important to clarify that legislative instruments in the form of a regulation are binding on all MS of the EU and do not need to be transposed as they enter into force on domestic legal systems from the moment that they are adopted by the EU legislators. The EU only resorts to regulations in frameworks that are of high relevance to the fulfilment of the EU general principles and when it is politically feasible to require harmonization and standard application throughout the whole EU.

tionally, the DSA places high attention on commitments already agreed in self-regulation practices, such as the Code of Conduct on countering illegal hate speech online.[189] In a complementary manner, the EU AI Act, also addresses online harms by establishing procedural safeguards. Still, both instruments focus on the procedural law rather than the substantive definitions of illegal content, thus not defining hate speech.

Second, in 2024, the European Parliament and the Council adopted the Directive on combating violence against women and domestic violence,[190] which criminalizes cyber stalking; cyber harassment; and cyber incitement to violence or hatred. The new rules in this Proposal strengthen the access to justice for victims and complement the DSA. As this legislative instrument in the form of a Directive, though binding on Member States, these have discretion in transposing it into the domestic legal system. Still, it certainly brings legal cohesion on the regulation of hate speech on the grounds of gender and sex.

Finally, the EC also recently expressed in December 2021 its intent to extend the list of the EU crimes to hate speech and hate crime given its growing impact, especially with technological changes, and given the lack of coherence in the criminalization of hate speech and hate crime among Member States.[191] This initiative aligns with the DSA and the Directive combating violence against women and domestic violence. The collection of these two legislative initiatives would add to the *lex specialis* on hate speech and potentially amend the discussed shortcomings in the Framework Decision.[192] The EC has argued in favour of European standardization stating that hate speech: contains a cross-border dimension; is a relevant area of crime particularly serious as it undermines EU fundamental rights in Art. 2 and 6 of the TEU and the CFR; presents recent worrying developments, especially with online hate speech. The Council of the EU and the EP now have to agree that these areas represent another area of crime in need of standardization across the EU.

189 European Union, Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, Article 93.

190 European Union, Directive (EU) 2024/1385 of the European Parliament and the Council of 14 May 2024 on combating violence against women and domestic violence, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401385> accessed 20 November 2024; *see also* European Commission Press Release IP/22/1533, International Women's Day 2022: Commission Proposes EU-Wide Rules to Combat Violence Against Women and Domestic Violence (Mar. 8, 2022).

191 Communication from the Commission to the European Parliament and the Council, COM (2021) 777 final (Sept. 12, 2021); *see also* European Commission Press Release IP/21/6561, The Commission proposes to extend the list of 'EU Crimes' to hate speech and hate crime. (Dec. 9, 2021).

192 *Id.*

2.4.3 Soft Law

The EU Code of Conduct on countering illegal hate speech online (Code) was adopted in 2016 and is a non-binding self-regulatory instrument.[193] The Code does not define hate speech. Instead, it adopts the definition of "illegal hate speech" of the Framework Decision, which only grants protection on the grounds of race, colour, religion, descent or national ethnic origin. There are two problematic aspects with this framework.

First, the distinction between legal and illegal hate speech is problematic as the utilization of *illegal* hate speech seems to imply there is legal hate speech. It is true the term hate speech has been *informally* used to refer to expressions regulated as criminal offenses, civil or administrative breaches, or even to harmful expressions not comporting legal consequences. Nevertheless, the ECtHR has only referred to hate speech to cover criminal offenses or civil or administrative breaches. Therefore, the third category of expressions, i.e., harmful expressions without legal implications, should not be considered hate speech at all. The utilization of legal hate speech for this third type of expressions induces legal unclarity and enforceability which compromises the regulatory efforts of regulating hate speech.

Second, the restriction of the impermissible grounds of hate speech to accounts of race or xenophobia is not aligned with the case law of the ECtHR and with the open-ended list of grounds protected from discrimination under Article 21 of the CRFEU.

Some remarks are necessary with regard to the procedural requirements stemming from the Code. It is confusing to note that despite pointing in the direction of a definition of hate speech, the Code prescribes that the IT companies will nevertheless review the illegality of the content against their community guidelines and not against the Framework Decision or national laws transposing it. Additionally, it is concerning that the Code gives the lead in this balancing of rights to the IT Companies without instructing them on the broader European human rights law and interpretative case law. The latter remark is particularly important in the case of IT companies of substantive size and thereby potentially delivering services comparable to public service. Moreover, the Code does not require reporting on the methodology for monitoring practices of hate speech which renders any study on the efficacy of the measures impossible and therefore does not comply with the conditions for restriction of freedom of expression as per Article 10(2), i.e. in that it is important to prove that the restriction is prescribed by law, in pursuit of legitimate interests and necessary in a democratic society.

---

193 *European Union Code of Conduct on Countering Illegal Hate Speech Online*, EUROPEAN COMM., https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discriminati on/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech -online_en.

2.4.4 Overview European Union, Hate Speech and Historical Oppressions

In summary, the legal strategies of the EU to conceptualize hate speech fall short from formally aligning with the original conceptualization by Matsuda[194] of hate speech as expressions perpetuating historical, systematic and intersectional systems of oppression.

The EU has in different moments adopted conflicting positions in regards to the intersectionality of systems of oppression. Although the AVMSD refers to the open-ended list of impermissible grounds for hate speech as per Article 21 of the CFR,[195] in the Framework Decision the EU limited the protection to people targeted on the basis of race, colour, religion, descent or national ethnic origin. This shows internal legal incoherence and a significant deviation from the intersectionality theory, which highlights the need to grant protection to any person from hateful expressions reinforcing a historically vertical relationship (among others, women, the queer community, and persons with disabilities).

2.5 Main Elements of Hate Speech in the European Context

The following Section describes the main legal elements of hate speech in the European context by building on the findings from its conceptualization at the Council of Europe and at the European Union. This exercise will consider in particular how such main elements of hate speech in the European context relate to the legal theoretical foundations on hate speech stemming from critical legal theory, specifically from critical race theory. It is important to reiterate that this concluding Section in the Article does not purport to reach a clear definition of what hate speech is. Instead, the main goal is to critically distil the main elements surfacing in the European context about the regulation of hate speech.

2.5.1 What Is Hate Speech and What Does It Do?

Although there is no agreed, legally binding definition of hate speech at the European level,[196] the findings from the analysis of the instruments at the Council of Europe and at the European Union level help present what the

194 The conceptualization of (racist) hate speech introduced by Matsuda in Section 2 of this Chapter.

195 European Union, *Charter of Fundamental Rights of the European Union*, 26 October 2012, 2012/C 326/02, https://www.refworld.org/docid/3ae6b3b70.html .

196 Neither at the European nor at the international level.

general common understanding of hate speech is per the European human rights standards.

Hate speech can be broadly conceptualized as any kind of expression that incites, promotes, spreads or justifies violence, hatred or discrimination against one person or a group of people based on presumed or real identity characteristics such as race, religion, sex, gender, sexual orientation, citizenship, national or ethnic origin, age, or disability. Four explanatory remarks are due.

First, the meaning of "expressions" should be construed in a broad manner and be understood as *any* expression be it verbal, non-verbal (such as facial expressions, gestures, pictures, signals), shared in person or online (including for example GIFs and memes). This is supported by CM/Rec(2022)16 when it prescribes that hate speech includes "all types of expressions." Furthermore, this is also the understanding in the Istanbul Convention when in Article 40 it includes "verbal and non-verbal expression."

Second, the impact of the expression should also be construed in an expansive way to include dissemination, promotion, and support. This is supported by CM/Rec(2022)16 when in paragraph 1(2) it refers to expressions that "incite, promote, spread or justify violence, hatred or discrimination."[197] Moreover, this is also aligned with the inclusion of dissemination as a qualifying act of hate speech by APCC Article 3 and ECRI GPR 6 as well as with the inclusion of aiding and abetting contained in the Istanbul Convention Article 41.

Third, the list of identity characteristics protected from hate speech is to be broadly conceptualized, too, and to be considered open-ended. The ECtHR, to date, in its interpretation of the ECHR has ruled on cases of racial, ethnic, religious, and homophobic hate speech. In its case law, there is no indication of limits regarding the impermissible grounds of hate speech. Furthermore, taking as the point of departure human rights provisions on non-discrimination, both Article 14 ECHR and Article 21 of the CFREU include "such as" before illustrating the list of impermissible grounds for discrimination. However, some confusion in the application of such a standard may arise when looking into some sector specific regulatory instruments on the criminalization of hate speech both at the EU and at the CoE level.

For example, within the CoE regulatory framework, there are distinctions in the protection granted from hate speech by sector specific treaties. For example, while the APCC only protects against racism and xenophobic speech, the case of the Istanbul Convention is more complex. Though the latter stresses the need to protect people from hate speech on the basis of their sex and gender, the drafters of the Explanatory report suggested the adoption of an open-ended list for grounds for non-discrimination, including gender, sexual orientation, gender identity, age, state of health, disability, marital status, and

197 U.N. Doc. A/HRC/22/17/Add.4 (n 70).

migrant or refugee status.[198] As a result, with this expansive interpretation of the groups protected under the Istanbul Convention from hate speech, the CoE currently grants redress through criminal law to an open-ended list of impermissible grounds.

Within the EU legislative framework, there seems to be a disconnect between the protected categories in the Framework Decision and in the Code of Conduct (which aligns with the Framework Decision) when compared with the conceptualization of protected categories under the AVMSD. To clarify, on one hand, the AVMSD aligns with the CFREU and addresses hateful expressions on the grounds of a wide range of categories including sexual orientation, disability, and age. On the other hand, the Framework Decision and the Code of Conduct address only racist and xenophobic hate speech. Still, it should be highlighted that, for the purposes of criminalization, the AVMSD refers to the hate speech conceptualization of the Framework Decision. As a result, the EU currently only grants access to criminal law to people targeted with hate speech on the grounds of their race, colour, religion, descent, or national or ethnic origin.

As observed by Alkiviadou more generally and emphasized here in the context of the EU, the European legal strategies can be interpreted as validating a "hierarchy of hate,"[199] where only certain forms of hatred are classified as impermissible and only certain affected groups could seek redress under criminal law. This is a point of concern, particularly in the EU system, because it overlooks research on enhanced prevalence of hatred based on victims' intersectional characteristics and because recent data shows that online hate speech across the EU level is mostly registered toward group's identities on the basis of sexual orientation.[200] Furthermore, these legal strategies separating systems of oppression fail to address the works from critical legal and race theorists when they were alerted to the intersectionality of socio-historical contexts of oppression (i.e., how someone targeted by different systems of oppression would need better protection from hate speech; an example of someone targeted by different systems of oppression could be someone racialized, a woman, a queer individual, or a disabled individual). As a result, the only route to progress toward a more coherent and cohesive European human rights framework is to expressly adopt an open-list for grounds prohibiting discrimination, whereby intersectionality of systems of oppression is clearly addressed as a criterion in the legal assessment.[201]

198 Convention on Preventing and Combating Violence Against Women and Domestic Violence (n 130), ¶ 53.

199 Natalie Alkiviadou, *The Legal Regulation of Hate Speech: The International and European Frameworks*, 55 Politièka Misao 203, 223 (2018).

200 Didier Reynders, Directorate-General for Justice and Consumers, European Commission, 5th valuation of the Code of Conduct on Countering Illegal Hate Speech Online (June 4, 2020).

201 Though important to continue to guard against the dilution of grounds.

The fourth and final remark on what hate speech is and what impact it has on its targets relates precisely with the fact that hate speech, as per its foundational conception, targets people who exist in systems of oppression. It is worth reminding that, the original conception in critical legal theory presented by Matsuda, "the hateful message is directed against a historically oppressed group and reinforces a historically vertical relationship."[202] This is a relevant reminder, especially with the current interpretation that the list of impermissible grounds for hate speech is open-ended. The interplay between these two conceptual elements should be translated to mean that, although any person can be targeted by hate speech based on a given group categorization, hate speech primarily functions as a means to keep the targeted group oppressed. A caveat should nevertheless be introduced to grant protection from hate speech in case of a "pressing social need" to protect national security, territorial integrity or public safety, prevention of disorder or crime, for the protection of health or morals, reputation or rights of others, prevention of the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

With regard to the legal framework for measuring the existence of systems of historical or systematic oppression, this Chapter suggests focusing on analysing the group's representation in political, economic, and social spheres. For instance, if the group has effectively exercises their rights to education, to work, or to access information. In contexts where policies of affirmative action have been designed and implemented, these represent a good framework for examining whether the group endures systematic oppression.

For clarity, this Chapter gives two examples of the application of this framework. For instance, white, heteronormative, cisgender, neurotypical, and able men, in principle belonging to various privileged societal groups, would not be immediately granted protection under hate speech laws unless proven that, with the experience of being targets of such speech, the protection of the group from such speech had become a pressing social need as per the ECtHR case law. This reasoning can be applied to the real case scenario mentioned in the introduction Section of this Chapter. To recapitulate, in the wake of a terrorist attack in London in 2017, a U.S. congressman wrote a Facebook post in which he directly called for the slaughter of all "radicalized" Muslims. This post pleading for imminent violence went untouched by Facebook. Later that year, a post on Facebook by a Black Lives Matter activist saying "All white people are racist" drew a different response. This post was removed, and the account was disabled for seven days. Facebook justified the different responses by stating "radicalized Muslims" was a sub-group of protected groups, while "all white people" was a more encompassing group and therefore deemed it more critical to protect. To properly address such a case, the current European human rights standards would need to have an explicit reference to the

202 Matsuda et al. (n 20), 35.

balancing between systems of oppression in order to deliver on values of democracy and human rights, including the rights of minorities.

Another example is the high levels of hate speech directed at health professionals during the COVID-19 pandemic. Here, although there is not an historical oppression of people belonging to such a group, the high levels of systematic threats had a substantial impact on how such professionals could live their lives and, in this case, they would certainly be protected under hate speech laws.

2.5.2 How to Substantively Regulate Hate Speech

The below Section introduces five main axes or principles for the regulation of hate speech: (i) hate speech is always illegal speech; (ii) hate speech laws aim to protect first and foremost historically oppressed groups; (iii) hate speech is only criminalized in its most serious forms; (iv) if not criminalized, then balancing of rights ideally on a case-by-case basis; and (v) regulating hate speech cannot simply be case-by-case.

2.5.2.1 Hate Speech Is Always Criminal or Illegal Speech

Hate speech should only be used to refer to either criminal offenses or problematic expressions, which although not criminal offenses should be prohibited under Article 10(2) of the ECHR and be legally actionable under civil or administrative law. Applying this framework to the ECHR, the term hate speech should only be used for expressions comprising an abuse of rights (Art. 17 ECHR) and for expressions, which despite being within the remit of freedom of expression (Art. 10(2) ECHR) can or should be legally limited in line with conditions explained in Article 10(2). The legal framework should not use the expression illegal hate speech as it induces the reader into thinking there is legal hate speech, therefore creating challenges of legal clarification and foreseeability. Expressions of intolerance and disrespect not amounting to hate speech but still raising issues of respect and tolerance should be addressed through culture and education.

2.5.2.2 People Targeted By Hate Speech Have Been Historically or Systematically Oppressed

The current European human rights standards need to explicitly acknowledge the linkage between hate speech and the perpetuation of systems of oppression. This is a key acknowledgement to promote the values of democracy and human rights, including the rights of minorities, guiding the European human rights framework. In considering the social background, the ECtHR should

expressly investigate socio-historical systems of oppression.[203] This also aligns with the "positive obligations doctrine" whereby States have a positive obligation to investigate bias indicators (i.e., objective facts or circumstances by which probable motives can be discerned).[204] The necessity test must also consider the victims' perspectives. In looking into the victims' perspective, the relevance of the intersectionality framework must be highlighted to better reflect the "living instrument doctrine," i.e., to have interpretations that adequately reflect present-day realities of targeted groups. The list of impermissible grounds for hateful expressions must be open-ended to account for present day realities of oppression (e.g. queerphobic speech, hate targeting refugees).

2.5.2.3 Hate Speech Should Only Be Criminalized In Its Most Serious Forms

European human rights standards require that only the most serious forms of hate speech be criminalized. The ECtHR adopts an assessment on a case-by-case basis. Nevertheless, its case law under Article 17 should guide the European framework regulating hate speech amounting to criminal offences. The jurisprudence of the ECtHR regarding expressions that constitute the most serious cases of hate speech was summarized in Paragraph 11 of the Appendix to CM/Rec(2022)16 and it includes:

a. public incitement to commit genocide, violence or discrimination;
b. racist, xenophobic, sexist and LGBTI-phobic threats;
c. racist xenophobic, sexist and LGBTI-phobic public insults under conditions such as those set out specifically for online insults in the Additional Protocol to the convention on Cybercrime concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems (ETS No. 189);
d. public denial, trivialization and condoning of genocide, crimes against humanity or war crimes; and,
e. intentional dissemination of material that contains such expressions of hate speech (listed in a-e above) including ideas based on racial superiority or hatred.[205]

Notwithstanding this useful summary of the most serious cases of hate speech contained in CM/Rec(2022)16, the list of impermissible grounds for hate speech should not be strictly interpreted but rather open-ended. That is because it is relevant to investigate the history and intersectionality of systems of oppression in the given context where the hateful expression was communi-

203 Sahana Udupa, *Decoloniality and Extreme Speech*, Paper Presented at the 65th e-Seminar, Media Anthropology Network, European Association of Social Anthropologists 24 (2020).
204 *See, e.g.*, Ibentoba v. Georgia, App. No. 73235/12 (May 12, 2015), https://hudoc.echr.coe.int/fre?i=001-154400.
205 U.N. Doc. A/HRC/22/17/Add.4 (n 70).

cated (as just explained in Section 2.5.2.2). The victims' perspective must be construed to allow for an interpretation that is reflective of present-day realities (e.g. queerphobic speech, hate targeting refugees).

The limited classification of criminal hate speech only for the most serious forms of hate speech enables the consequential application of the right to an effective remedy as prescribed in Article 13 of the ECHR. All other forms of hate speech, illegal but not criminal, can lead to administrative or civil action, thus avoiding legal uncertainty on criminal action and thereby contributing to a more effective right to remedy.

2.5.2.4 If Not Criminalized, Then Need to Balance Human Rights

"Freedom of expression is the condition *sine qua non* for a genuine pluralist democracy":[206] The ECtHR ruled in Handyside v. UK that expressions that "offend, shock or disturb" are generally protected by the ECHR, considering that "such are the demands of pluralism, tolerance, and broadmindedness without which there is no democratic society."[207] States have to justify even minor interferences with Article 10.[208] Expressions that are prohibited but not criminal (thus not triggering the application of Article 17 ECHR) must respect conditions as per Article 10(2) ECHR: provided by law; pursuing one of the mentioned legitimate purposes; and, be necessary in a democratic and pluralistic society, following the ECtHR jurisprudence. The necessity test must encompass the analysis of the following contextual factors: political and social background; intent of the speaker; speaker's status or role in society; content of the expression; extent of the expression; and the nature of the audience. In this exercise of assessing whether a limitation on the right to freedom of expression is necessary in a democratic society, it is important to explicitly account for socio-historical records of oppression and the victims' potential intersectional position between various oppressive systems.

2.5.2.5 Challenges With the Regulation of Online Hate Speech

New technologies, social media and algorithms designed to promote content virality have greatly contributed to the increase of online hate speech. Given the prevalence of hate speech online, it is impossible to assess posts on a case-by-case basis. Regulation of online harms in general, and of hate speech in

206 Françoise Tulkens, *When to Say Is To Do, Freedom of Expression and Hate Speech in the Case-Law of the European Court of Human Rights*, Seminar on Human Rights for European Judicial Trainers (Sept. 3, 2013). Tulkens defends that freedom of expression is both a safeguard against interference by the State (subjective right) and an objective right, and a means for the establishment of a democratic society.

207 Handyside v. UK (n 102).

208 *See, e.g.*, Ottan v France, App. No. 41841/12 (July 19, 2018), https://hudoc.echr.coe.int/fre?i=001-182627.

particular, focuses on instructing internet intermediaries on minimum human rights procedural safeguards and public monitoring and auditing of the digital platforms' compliance with due diligence frameworks. Internet intermediaries often moderate online content with the use of automated decision-making tools. Internet intermediaries should design and employ automated-decision making tools to identify, ex ante, at least the most serious cases of hate speech, as elaborated in the jurisprudence of the European Court of Human Rights on Article 17 of the European Convention on Human Rights.

2.6 Conclusion

Scholars, practitioners, and policy-makers have long focused on clarifying the definition and status of hate speech in international and regional human rights. However, this has proven to be a challenging process and the absence of a legally-binding definition of hate speech in human rights law has had severe negative individual and societal implications. In an era of digital communication where there is an increased prevalence and reach of hate speech, it is imperative to advance a standardized legal conceptualization of hate speech that is suitable to protect and to present legal remedies for people targeted by such hateful expressions.

This Chapter clarifies the original conceptualization of hate speech advocated by critical race scholars, grounded on the perpetuation of intersectional, historical or systematic oppression. This Chapter then analyses, in the light of that original conceptualization of hate speech, a selection of legal initiatives in the European context, covering treaties and non-treaty initiatives suggested as the most relevant in the regulation of hate speech at the European level. The main treaty instruments analysed are the European Convention on Human Rights and the Charter of Fundamental Rights of the EU. The most relevant EU legal instruments are the Framework Decision on Combating Racism and Xenophobia, the Audiovisual Media Services Directive and the Code of Conduct on countering illegal hate speech online. The main non-treaty instrument is the Recommendation of the Committee of Ministers of the Council of Europe on a comprehensive approach to hate speech (CM/Rec/(2022)16).

The analysis in this Chapter explains the interplay between European regulatory instruments and claims that a more standardized conceptualization of hate speech rooted in the intersectional, historical or systematic systems of oppression perpetuated by hate speech can help reconcile the regulation of hate speech in Europe. If the European regulatory framework to counter hate speech is to uphold values of equality, dignity, pluralism, and democracy, then the most effective and legally coherent manner to achieve that objective is to emphasize the seminal hate speech elements presented by critical race and legal scholars and to emphasize the need to investigate intersectional, historical or systematic systems of oppression perpetuated by hate speech.