Ca' Foscari University of Venice

Master's Degree programme

in Comparative International Relations

Final Thesis

Hate speech in international law

Searching for the complex balance with freedom of expression

Supervisor
Ch. Prof. Sara De Vido

Graduand
Laura Barison
Matriculation number: 882989

Academic Year
2020 / 2021

Abstract p. 3

Introduction p. 7

Chapter 1: Hate Speech Law

1.1 What is Hate Speech p. 12
1.1.1 The origin of Hate Speech: Discrimination on the basis of Gender, Race, Religion and Sexual Orientation p. 14
1.1.2 A Competing Right: Freedom of Speech and its definition p. 17
1.2 Hate Speech in International Law p. 19
1.2.1 Universal Declaration of Human Rights p. 19
1.2.2 International Covenant on Civil and Political Rights p. 21
1.2.3 Convention for the Elimination of Racial Discrimination p. 25
1.2.4 Relevant UN documents: the Rabat plan of action and the reports of the Special Rapporteur on Freedom of Opinion p. 28
1.3 Regional Hate Speech Law and Approaches p. 33
1.3.1 European Hate Speech Law: Council of Europe and the European Union p. 34
1.3.2 Interamerican Commission on Human Rights Hate Speech Law p. 39
1.4 European Court of Human Rights Jurisprudence on Hate Speech p. 41
1.4.1 Inconsistencies in the Jurisprudence p. 46
1.5 Interamerican Court of Human Rights Jurisprudence on Hate Speech p. 48

Chapter 2: Controversies in Hate Speech Regulation

2.1 Freedom of Speech and Hate Speech: Balancing approach problems p. 51
2.1.1 The absence of a broadly accepted definition and clear standards for Hate Speech p. 51
2.1.2 The ideological debate: is the limitation to free speech acceptable? p. 55
2.2 Online Hate Speech p. 60
2.2.1 Characteristics of online hate speech p. 62
2.2.3 Multilateral regulation for online hate speech p. 66
2.2.3 The responsibility of private companies p. 69
2.3 Biases in Hate Speech regulation p. 72

2.3.1 Protection from discrimination and hate speech on the basis of sex p. 73
3.2.2 Sexist hate speech online p. 79
2.3.3 The protection of women and the LGBTQ community under hate speech laws: theory vs practice p. 85

Chapter 3: Different national approaches to the regulation of Hate Speech

3.1 Germany: the legal framework on hate speech and the NetzDG p. 89
3.1.1 The German legal framework for hate speech p. 90
3.1.2 The NetzDG: Combating hate speech on social media p. 92
3.2 France: Hate Speech laws and the failure of the "Avia Bill" p. 94
3.2.1 Overview of the French provisions on discrimination and Hate Speech p. 95
3.2.2 The "Avia Bill" on online Hate Speech: criticisms and concerns p. 99
3.2.3 The decision of the French Constitutional Council p. 102
3.3 Italy and the limited scope of Hate Speech laws p. 104
3.3.1 Provisions directly and indirectly restricting hate speech p. 106
3.3.2 The role of equality institutions in Italy p. 110
3.3.3 Protection of LGBTQ people: the rise and fall of DDL Zan p. 113
3.4 United States' liberal vision on Hate Speech: The First Amendment p. 116
3.4.1 The First Amendment: history and first interpretations p. 116
3.4.2 US Supreme Court Jurisprudence p. 119
3.4.3 Snyder v. Phelps, a landmark decision on Freedom of Speech p. 123

Conclusion p. 127

Bibliography p. 130

Abstract

Questo lavoro è nato con l'obiettivo di fornire un contesto alle normative esistenti a livello internazionale nell'ambito del cosiddetto discorso d'odio (hate speech in inglese) e di chiarire il framework in cui si sviluppano gli strumenti volti a limitare il diritto alla libertà di espressione nei casi in cui essa vada a ledere la dignità di altri individui e ad incitare alla violenza. Inoltre, vengono evidenziate lungo il percorso sia le diverse difficoltà incontrate nell'applicazione delle principali norme internazionali sia le caratteristiche peculiari del discorso d'odio, con il fine di dare concretezza a un concetto non privo di ambiguità.

Il lavoro si articola in tre capitoli, completi di introduzione e conclusione. All'interno del primo capitolo vengono analizzate le principali norme applicabili ai casi di discorso d'odio sia a livello internazionale che regionale. Innanzitutto, viene chiarito che le normative internazionali sono sostenute da due differenti gruppi di provvedimenti, quelli antidiscriminazione e quelli atti a garantire la libertà di espressione. I primi sono la risposta alle ineguaglianze presenti nella società e sono il primo strumento che può essere utilizzato nel contrastare il fenomeno, in quanto il discorso d'odio può essere definito come un attacco basato sulle caratteristiche personali di un individuo o di un gruppo. Nonostante la crescente applicazione di questo genere di norme negli ultimi decenni è evidente come la loro attuale forma non sia sufficiente per contrastare tutte le nuove forme di discriminazione, in particolare per quanto riguarda un fenomeno relativamente nuovo quale il discorso d'odio. Oltre alle norme antidiscriminazione, il corpus di normative internazionali applicabili ai casi di discorso d'odio comprende tutte le norme relative alla libertà di espressione. Questa è indubbiamente uno dei diritti fondamentali della persona, essenziale per lo sviluppo delle caratteristiche individuali di ciascuno e per il corretto funzionamento di uno stato democratico. Nonostante ciò, la libertà di espressione non va intesa, almeno secondo quanto emerge dall'analisi delle normative internazionali vigenti, come un diritto assoluto. Infatti, tramite lo studio delle convenzioni internazionali entrate in vigore poco dopo la Seconda Guerra Mondiale, è possibile individuare un certo numero di clausole atte a limitare, in casi di abuso di diritto o di rischio per la sicurezza dello stato e della comunità, il diritto alla libertà di espressione. Tali limitazioni sono il centro della normativa internazionale in materia di discorso d'odio, in quanto il concetto è stato teorizzato solo negli anni '80 del Novecento e pertanto il termine non viene utilizzato se non in documenti più recenti. Nella Dichiarazione Universale dei Diritti Umani tale limitazione risulta implicita, identificabile tramite l'analisi congiunta degli articoli 7 e 29, riferiti l'uno al principio di eguaglianza e non discriminazione, e l'altro alle possibili limitazioni dei diritti enunciati nella Dichiarazione "per assicurare il riconoscimento e il rispetto dei diritti e delle libertà degli altri e per

soddisfare le giuste esigenze della morale, dell'ordine pubblico e del benessere generale in una società democratica". Nel successivo Patto Internazionale Relativo ai Diritti Civili e Politici delle Nazioni Unite, la connessione fra il diritto alla libertà di espressione e la sua limitazione diventa esplicita; infatti, l'articolo relativo alla libertà di espressione enuncia al paragrafo 3 le limitazioni a cui tale diritto può essere sottoposto, per il rispetto dei diritti e della reputazione altrui e per la protezione della sicurezza nazionale, dell'ordine pubblico, della sanità o della morale pubblica. La Convenzione sull'Eliminazione di tutte le Forme di Discriminazione Razziale del 1969 fornisce una garanzia ulteriore in materia di discorso d'odio, in quanto richiede ai firmatari di condannare ogni forma di discriminazione razziale dichiarando come crimini punibili secondo la legge la diffusione di idee basate sulla superiorità o sull'odio razziale, l'incitamento alla discriminazione e l'incitamento alla violenza.

Per ottenere un quadro completo della normativa internazionale sul discorso d'odio è necessario analizzare, oltre alle principali convenzioni internazionali, anche l'approccio delle Nazioni Unite nei confronti del fenomeno e le normative regionali in materia. Per quanto riguarda le Nazioni Unite è importante citare il Piano d'Azione di Rabat riguardante le norme contro l'incitamento all'odio, nel quale viene reiterata la necessità di limitare il diritto alla libertà di espressione nei casi di incitamento. Inoltre, anche il Relatore Speciale delle Nazioni Unite per la libertà di espressione si è occupato di discorso d'odio, fornendo in uno dei suoi rapporti un importante test per la valutazione dell'espressione applicabile ai casi di incitamento all'odio. A livello regionale invece, l'approccio di due importanti organizzazioni europee aiuta a ricostruire il framework della regione. In primo luogo, il Consiglio d'Europa ha pubblicato varie raccomandazioni relative alle limitazioni della libertà di espressione, rinforzando la nozione di non assolutezza del diritto. Inoltre, la legislazione dell'Unione Europea richiede esplicitamente la criminalizzazione del discorso d'odio razzista e xenofobo. Questo approccio si può individuare anche nella giurisprudenza della Corte Europea dei Diritti dell'Uomo, la quale ha giudicato 60 casi di discorso d'odio dalla sua fondazione. Al contrario, nel contesto interamericano, le norme relative a questo tipo di limitazioni sono poche e non vincolanti. Infatti, nonostante la Commissione Interamericana dei Diritti Umani abbia confermato la presenza di limitazioni per il diritto alla libertà di espressione, non è stato possibile creare legislazioni vincolanti a riguardo, probabilmente a causa della forte tradizione garantista statunitense.

Nel secondo capitolo il focus viene spostato dalla normativa internazionale alle difficoltà riscontrabili nell'applicazione pratica della normativa e alle caratteristiche peculiari del fenomeno. In primo luogo, va evidenziato come non esista un vero e proprio consenso su cosa sia il discorso d'odio. A livello internazionale le Nazioni Unite hanno cercato di colmare questo vuoto fornendo una definizione nel contesto della strategia sull'Hate Speech del 2020. Nonostante ciò, non vi è consenso tra stati, studiosi

e politici su quale sia la definizione effettiva di questo concetto. Per questo motivo, l'applicazione legale del concetto risulta difficile. Inoltre, l'inabilità di individuare le caratteristiche specifiche e i limiti di applicazione che le normative sul discorso d'odio dovrebbero avere porta ad una inconsistenza negli standard di applicazione delle norme internazionali. Inoltre, su un piano ideologico, lo stesso concetto viene sottoposto a severe critiche. Secondo alcuni studiosi, l'utilizzo di una parola come "odio" per definire un tipo di espressione vietata dalla legge è scorretto, in quanto il collegamento tra un sentimento quale l'odio e l'espressione può portare ad una visione limitante delle condizioni in cui applicare il concetto. In aggiunta, la difficoltà nell'identificare i limiti entro cui l'applicazione del concetto si innesta viene vista come un rischio, in quanto la limitazione del discorso d'odio potrebbe essere usata come giustificazione per limitare forme legali di espressione. La stessa limitazione della libertà di espressione è tuttora soggetta a scrutinio. Nonostante le norme internazionali la prevedano, la tradizione liberale basata sulla filosofia di Mill del "libero mercato delle idee" e le caratteristiche essenziali della libertà di espressione per lo sviluppo della persona portano alcuni a considerarla inviolabile.

Oltre alle difficoltà che comporta la mancanza di consenso riguardo alla definizione del fenomeno, viene analizzata la questione del discorso d'odio online. Diffusosi ampiamente negli ultimi anni, Internet ha portato con sé una serie di cambiamenti sconvolgenti nella vita di tutti i giorni. Alcune caratteristiche della rete si sono dimostrate pericolosamente utili alla diffusione del discorso d'odio online. In particolare, l'anonimità, l'immediatezza, la semplicità nel trovare una comunità e la portata internazionale fornite dalla rete hanno facilitato e fomentato la diffusione di questo fenomeno. Per contrastarlo, date le caratteristiche specifiche di Internet, risulterebbe necessaria una collaborazione multilaterale tra stati, non raggiunta finora. Inoltre, la componente privata della rete aggiunge un ulteriore livello di difficoltà alla lotta al discorso d'odio online, in quanto risulta difficile individuare i responsabili della diffusione dei contenuti.

Per concludere lo studio delle principali sfide riscontrabili in ambito di regolamentazione del discorso d'odio, il secondo capitolo fornisce un'analisi dei principali bias riscontrabili nell'applicazione delle norme. In particolare, viene analizzato il fenomeno del discorso d'odio sessista o sulla base del genere, tanto diffuso quanto spesso ignorato dalle legislazioni e dalle corti internazionali. Infatti, il discorso d'odio sessista, diffusissimo soprattutto sui social network, risulta essere generalmente poco citato dalle normative in materia, e di conseguenza pochi sono i casi in cui le corti si trovano a giudicare su questa branca del discorso d'odio. Similmente, è rara l'applicazione di tali norme per i casi di discorso d'odio relativi all'orientamento sessuale o al genere, altro fenomeno molto diffuso sia sui social network che nella vita di tutti i giorni.

Nel terzo ed ultimo capitolo, l'elaborato si propone di fornire degli esempi nazionali di normative in materia di discorso d'odio, rappresentativi delle difficoltà e delle criticità precedentemente analizzate. In primo luogo, viene analizzata la normativa tedesca, considerata una delle più complete nel contesto europeo e mondiale, anche a causa delle particolari condizioni storiche nella quale si innesta. Dopo l'esperienza nazista degli anni '40 del Novecento, infatti, la Germania ha istituito un corpus di leggi che limitano la libertà di espressione nei casi di incitamento all'odio verso parti della popolazione, assalto alla dignità della persona o incitamento di masse. Nel 2017 la Germania è stato il primo paese europeo ad introdurre una normativa, la cosiddetta legge NetzDG, in materia di discorso d'odio online, riguardante la responsabilità delle compagnie private nella diffusione dei contenuti. Fortemente criticata, la normativa è stata usata come ispirazione da un altro paese europeo analizzato nel terzo capitolo, la Francia. In Francia, nel 2019, si è tentato di approvare una legge simile, la legge Avia, la quale è stata però ritenuta incostituzionale dalla Corte di Cassazione francese. Il terzo caso analizzato è quello italiano, nel quale si evidenziano le mancanze riscontrabili nella attuale legge antidiscriminazione, la legge Mancino del 1993, comprendente solo la discriminazione sulla base della nazionalità, dell'etnia e del credo religioso. Viene inoltre analizzato il travagliato percorso del DDL Zan, proposta di legge che mirava alla modifica della legge precedentemente citata, bloccato dal senato lo scorso ottobre. Il terzo capitolo si conclude con l'analisi dell'approccio statunitense al diritto alla libertà di espressione e alle sue, molto rade, limitazioni, tramite l'individuazione dei principali strumenti utilizzati dalla Corte Suprema nei casi di discorso d'odio e l'analisi di un caso specifico, Snyder contro Phelps.

Introduction

The United Nations defines hate speech as:

> any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.[1]

The definition was provided in 2020, in an attempt to respond to the recent surge in hate speech cases around the globe. According to a report of the Geneva International Centre for Justice, most western democracies have witnessed a rise in cases of hate speech and discrimination in the last few years.[2] At the beginning of 2021, the possible alarming and dangerous effects of unregulated extreme speech have been shown to the world with the storming of the United States Capital by supporters of Donald J. Trump.[3] In this instance, the insurrectionists had coordinated for weeks on social media, encouraged by the false narrative of the "stolen elections", and the online exchanges turned into violent action on the 6th of January.[4] In Europe, a survey carried out by The European Union Agency for Fundamental Rights in 2019 found that 42% of interviewed LGBTQ people in the union had faced discrimination in the previous year, and 11% reported being physically assaulted.[5] "*In the Tsuruhashi district of Osaka, Japan, a 14-year-old girl addresses the crowd, saying how much she despises Korean people and wishes she could kill them all*".[6] Although the rise in hate speech around the globe is evident, the absence of a definition and of clear standards for what hate speech is makes the gathering of data inconsistent and scarce.[7] The term hate speech was first utilized by legal scholars in the 80's as an explanatory concept to group together and identify cases and laws that referred to specific types of discriminatory or offensive speech. Thus, the legal application of the term only came ex post, which created a series of difficulties in its application. In fact, although the definition

---

[1] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml, p. 9.

[2] Futtner, N. Brusco, N. 2021, *Hate Speech is on the Rise – Report*, Geneva, Geneva International Centre for Justice, p. 1

[3] Britannica website: https://www.britannica.com/event/United-States-Capitol-attack-of-2021 [Accessed 3 February 2022]

[4] Futtner, N. Brusco, N. 2021, *Hate Speech is on the Rise – Report*, Geneva, Geneva International Centre for Justice, p. 8

[5] European Union Agency for Fundamental Rights website: https://fra.europa.eu/en/data-and-maps/2020/lgbti-survey-data-explorer [Accessed 25 January 2022]

[6] Carlson, C. R. 2021, *Hate Speech*, Cambridge, The MIT Press, p. 10

[7] ILGA Europe website: https://www.ilga-europe.org/what-we-do/our-advocacy-work/hate-crime-hate-speech [Accessed 3 February 2022]

provided by the UN unites the most common interpretations of the concept, there is no official consensus on what hate speech is, as the term is not defined by any international convention.

Nonetheless, the spreading of the phenomenon has pushed many international bodies and national states to concentrate on combating it. The Secretary General of the United Nations Antonio Guterres addressed the issue of hate speech, considered as one of the most dangerous predecessors of hate crimes and atrocities,[8] in 2020. The Secretary General underlined how this type of speech can result in a setback for peace in the world and focused on the role of social media and the Internet in the proliferation of the phenomenon.[9]

This work stems from a desire to comprehend this complex reality more clearly, firstly by providing an in-depth analysis of the existing framework and an examination of the main international and regional instruments that allow the limitation of the right to freedom of speech. Later, the focus will shift to the main difficulties that arise from the vagueness that, to this day, characterizes the concept and the major challenges that the legal application of the concept poses, both in ideological terms and in practical terms, with the discussion of three emblematic national frameworks of legislation on hate speech.

More specifically, chapter one of the work will center on the analysis of the main international law and regional law provisions applicable to hate speech cases. The international definition and application of the concept relies on two main groups of provisions: anti-discrimination provisions and freedom of speech provisions. The first group is based on the principle of equality, as anti-discrimination laws are the response to the inequalities that can be found in societies.[10] Considering that hate speech can be defined as an attack to a person or a group based on its personal characteristics, anti-discrimination laws are the first tool that can be utilized in contrasting the phenomenon. Even though the application of these norms has grown in the last decades, it became evident that these instruments are not sufficient to eliminate inequalities or to combat the relatively new phenomenon of hate speech.[11] Freedom of speech provisions are at the same time the central tool in the fight to hateful expression and the most used justification for the lack of regulation on the matter. The right to freedom of expression is considered essential in democratic states.[12] However, it is also recognized that the application of the right in broad terms can constitute a danger for the legitimacy and

[8] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml, p. 7.

[9] United Nations website: https://www.un.org/sg/en/content/sg/statement/2019-06-18/secretary-generals-remarks-the-launch-of-the-united-nations-strategy-and-plan-of-action-hate-speech-delivered [Accessed 15 November 2021].

[10] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 5

[11] Ibid. p. 154.

[12] Fredman, S. 2018. "Freedom of speech" in *Comparative Human Rights Law,* pp. 305 - 354. Oxford: Oxford University Press, p. 307.

democracy of a state.[13] As a result, the provisions on freedom of speech provide explicit or implicit clauses allowing for the limitation of the right. In international law it is possible to identify different instruments that are applicable to hate speech cases and that provide for such limitations to freedom of speech: The Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the International Convention on the Elimination of All Forms of Racial Discrimination, the Rabat Plan of Action and the recommendations by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. However, seen as the concept was theorized in the 1980s, the most important international conventions on human rights do not provide specific definitions or provisions on the matter, and refer mostly to the concept of incitement to violence or discrimination. At the regional level, the European framework of regulation is the most elaborate and complete and, in the European Union, provides an explicit obligation for member states to criminalize racist and xenophobic hate speech.[14] Conversely, the Interamerican framework is characterized by few provisions and instruments related to hate speech, as well as the lack of any specific obligation for states on the matter, possibly because of the well-established absolute interpretation of the right of the United States.[15] Lastly, the first chapter provides an analysis of the jurisprudence of the European Court of Human Rights, which judged 60 hate speech cases from its establishment, and of the Interamerican Court of Human Rights.[16]

Chapter 2 will focus on the major issues and controversies that arise from the concept of hate speech and its practical application. The absence of an internationally accepted definition for the concept creates serious inconsistencies in its legal application. Scholars have argued that the use of the term "hate" to define a type of prohibited speech under international law can limit its understanding and possible application of the provisions.[17] Moreover, the difficulty in identifying the scope of possible hate speech provisions can leave space for problematic convictions and cases of abuse by states, with hate speech laws used to justify the delegitimization of political or dissenting speech in the state.[18] Furthermore, the confusion that arises from the merging of the concept of hate speech with that of incitement, often noticeable in national legislations, often results in the overlooking of cases of hate speech "*expressed in a civil or reasonable way or hate speech which has some bearing on a political issue or issue of public concern*".[19] From an ideological point of view, the legitimacy of the provision

---

[13] Ibid. p. 309.
[14] Eur-lex website: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l33178 [Accessed 4 December 2021].
[15] Carlson, C. R. 2021, *Hate Speech*, Cambridge, The MIT Press, p. 126.
[16]The Future of Free Speech website: https://futurefreespeech.com/hate-speech-case-database/ [Accessed 20 January 2022]
[17] Heinze, E. 2016, *Hate Speech and Democratic Citizenship*, Oxford, Oxford University Press, p. 22.
[18] Mchangama, J. 2015, "THE PROBLEM WITH HATE SPEECH LAWS", *The Review of Faith & International Affairs*, vol. 13, p. 78
[19] Ibid. p. 303.

on hate speech is still largely debated, with some scholars underlying the necessity of bans on hate speech to combat its effects and scholars vehemently opposing such bans, deemed counterproductive.[20] Moving to another critical point, online hate speech has further enhanced the difficulty in addressing the phenomenon. The peculiar characteristics of the Internet, international reach, anonymity[21], sense of community,[22] and immediacy,[23] make the fight to hate speech even more difficult. Moreover, the lack of a real multilateral regulation on online hate speech and the difficulty in identifying the responsibilities of online service providers add even more obstacles to the fight. Lastly, the analysis will center on the main biases that can be identified in current hate speech regulation, which most often than not fails to be applied to hate speech on the basis of sex or gender,[24] and does not address some very evident issues, such as online sexist hate speech.

The third and final chapter will provide an insight on four national frameworks of regulation for hate speech. The German legislation is one of the most complete and restrictive for what concerns hate speech cases. Moreover, the county recently passed a controversial law, the NetzDG Act, which addresses one of the main issues in hate speech legislation, online hate speech, and makes Internet providers liable for the sharing of hate speech on their platforms.[25] Similarly, the French legal framework allows for limitations to freedom of speech in cases of extreme speech or apology of terrorism.[26] However, contrarily to the German case, the attempt to adopt a law limiting online hate speech and making providers liable for the publication of content failed, as the Avia bill on online hate speech was deemed unconstitutional by the French Constitutional Council.[27] In the Italian case, the analysis of the provisions will highlight the shortcomings of the Italian framework on hate speech and discrimination in general, as the current law only criminalizes hate speech on racial, ethnic and religious grounds, leaving out sex, gender, sexual orientation, disability and age.[28] To conclude, the discussion will center on the U.S. approach to hate speech. In the country, the general belief is that freedom of speech should not be limited, unless it is to avoid a clear and present danger.[29] Thus, the

---

[20] Heinze, E. 2016, *Hate Speech and Democratic Citizenship*, Oxford, Oxford University Press, p. 33

[21] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 298

[22]Ibid. p. 301

[23]Ibid. p. 304

[24] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union, p. 157

[25] German Law Archive website: https://germanlawarchive.iuscomp.org/?p=1245 [Accessed 5 February 2022]

[26] Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States,* p. 17

[27] Conseil constitutionnel [CC] [Constitutional Court] decision No. 2020-801 DC, June 18, 2020, Rec. France

[28] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 22

[29] Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/249/47/#tab-opinion-1928047 [Accessed 29 January 2022]

First Amendment, the constitutional provision on freedom of speech, awards its protection to virtually all types of expression, regardless of the potential content.[30]

[30] Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, pp. 396

Chapter 1
Hate Speech Law

1.1 What Is Hate Speech

Recently there has been a notable increase in hate speech around the globe, which carries grave implications as it is considered one of the most common precursors to mass atrocity crimes.[31] In particular, at the launch of the United Nations Strategy and Plan of Action on Hate Speech, the Secretary General Antonio Guterres referred to hate speech as "*an attack to tolerance, inclusion and diversity*" and as a phenomenon that "*can lay the foundation for violence, setting back the cause of peace*". [32] Moreover, he underlined the role of new media in the rapid diffusion of hate speech. The term was coined by legal scholars in the late 1980s. This first use of the term was a purposeful one, the legal scholars of the early 80s started to employ the term to generalize and classify a particular group of laws that were in some way related to discriminatory or offensive types of speech.[33] Soon the term started to be utilized in human rights courts, as well as in some domestic legislations, thus overcoming the purely scholarly domain and becoming a de facto legal concept. Although we have examples of this concept being directly employed by courts,[34] accompanied by a definition, there is no consensus on how it should be interpreted in a broader, universal way.

A common approach to the issue of the definition of the term "*seeks to analyze the term hate speech through body of laws that may not contain the exact term hate speech*".[35] By using this approach, the legal employment of the concept draws from an array of different legal documents, mostly conventions, that seek to regulate certain types of acts that amount to discrimination and hatred, broadening the scope of such acts or using the norms contained in them as a basis on which to build a more general framework of applicable laws on hate speech. However, at the same time, this kind of approach leaves room for some inconsistencies and critical points. First, the amount in which the Hate Speech case should contain feelings of attitude of actual hatred. Second, the identification of groups that should be protected from hate speech. Third, the nature of the speech, which according to

[31] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml, p. 7

[32] United Nations website: https://www.un.org/sg/en/content/sg/statement/2019-06-18/secretary-generals-remarks-the-launch-of-the-united-nations-strategy-and-plan-of-action-hate-speech-delivered [Accessed 15 November 2021]

[33] Brown, Alexander. "What Is Hate Speech? Part 1: The Myth of Hate." *Law and Philosophy*, vol. 36, no. 4, 2017, p. 424

[34] Global Freedom of Expression. 2021. *Sürek and Özdemir v. Turkey - Global Freedom of Expression*. [online] Available at: https://globalfreedomofexpression.columbia.edu/cases/surek-ozdemir-v-turkey/ [Accessed 13 November 2021]

[35] Brown, Alexander. "What Is Hate Speech? Part 1: The Myth of Hate." *Law and Philosophy*, vol. 36, no. 4, 2017, p. 435

different scholars is defined as insulting speech, offensive speech, hostile verbal abuse, speech that denigrate people and so forth. Lastly, the broadness of the definition of speech acts that should be included in the category. Of course, this different definitions and approaches imply a difficulty in individuating the actual cases of hate speech.[36]. Therefore, the definition of the term Hate Speech remains a highly debated issue as there is no unified legal definition of hate speech and the definitions provided by states or other international instrument are generally slightly different in scope. The UN plan of action tries to address this issue and has the goal of providing a unified framework for combatting the phenomenon in the context of the United Nations. In the case of the Plan of Action, the UN defines hate speech as:

> any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.[37]

This definition, provided for the purpose of creating a common basis for the implementation of the plan of action, requires three components. First, hate speech can be delivered in any form: written, spoken, online, offline and so on. Second, hate speech must include an attack or discrimination, it is based on intolerant or prejudiced ideas or on biased or bigoted opinions. Third, hate speech is considered as such when it refers to an identity factor, hence making a reference to the protected characteristics of persons. Although exhaustive, this is not a legal definition, and it is not universally recognized nor accepted.

In European case law the term "Hate Speech" is used to identify a series of situations that can be summarized as incitement to racial hatred, incitement to religious hatred and incitement to other types of hatred based on intolerance.[38] The Committee of Ministers of the Council of Europe, instead, defines it as "*covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti- Semitism or other forms of hatred based on intolerance*".[39] The common ground in the majority of definitions of the concept of Hate Speech is that the focus should not be put on hateful acts in themselves. Another central issue in the debate over hate speech relates to the legitimacy of the concept in itself. In fact, according to many scholars, the existence of the concept of hate speech

[36] Brown, Alexander. "What Is Hate Speech? Part 1: The Myth of Hate." *Law and Philosophy*, vol. 36, no. 4, 2017, pp. 436-438

[37] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml, p. 9

[38] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 4

[39] Ibid. p. 3

and its punishability amounts to a violation of the right to free speech. "*All 'hate speech' laws, no matter how they are drafted, inherently violate the emergency and viewpoint neutrality principles*"[40] writes Nadine Strossen, drawing from the common First Amendment interpretation, stating that the government must not prohibit the expression of an idea only because it is considered offensive or disagreeable. Some scholars are, instead, more moderated in their analysis of the matter based on the balancing of the two approaches. In this light, a commonly proposed solution is to constrain free speech (therefore enforce hate speech regulation) only in cases where "*speech harms are sufficiently similar to conduct harms*"[41], hence focusing on the tangible harm that results from a hate speech act while maintaining the free speech principle untouched. In general, the legal concept of hate speech expresses the will to regulate, and in extreme cases punish, those types of speech that, in some ways, push certain ideals, stereotypes, hateful notions out into the society, with all the controversies that such approach to speech can arise.

1.1.1 The origin of Hate Speech: Discrimination on the basis of gender, race, religion and sexual orientation

As stated in the previous paragraph, one of the main approaches to the regulation of hate speech in international, regional, and national laws is the analysis of the body of anti-discrimination laws. This body of laws draws from a very well-known and overall accepted concept: equality. Although the principle of equality before the law is now considered a central principle in most western societies, it is a relatively new concept that appeared in conjunction with the rise of mercantile capitalism and the loosening of feudal bonds.[42] The rise of this principle allowed discriminated groups to argue for equal treatments and rights by giving them the vocabulary that was needed for such argumentations. The importance of this overarching concept in anti-discrimination law is undeniable, however it is important to note that even such a broadly accepted concept leaves room for different interpretations and, consequently, controversies. Equality could be interpreted as treating likes alike, as a correction of maldistribution of power and capabilities, as the equalization of opportunities.[43] By applying one of these underlying conceptions, the scope and objective of anti-discrimination laws, or of any law based on the equality principle, would be drastically different.[44] Anti-discrimination laws are the response to inequalities and have been developed mostly in the last century in the international system though the adoption of numerous conventions and included in most constitutions and legal systems.

---

[40] Strossen, N. 2018, *HATE - Why We Should Resist It with Free Speech, Not Censorship,* Oxford, Oxford University Press, p. 103
[41] Goldberg, Erica. "Free Speech Consequentialism." *Columbia Law Review*, vol. 116, no. 3, 2016, pp. 687–756., p. 689
[42] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 5
[43] Ibid. p. 8, 14, 18
[44] Ibid. p. 2

Discrimination laws are heavily influenced by the social and political context in which they have developed, for example: in the United States the development of such legislation began with race, while in the European Union, the process started with nationality and gender. In fact, the ECHR, developed in the early postwar period, reflects the context in which it was elaborated and therefore it includes birth, political opinion, and property but does not include, for example, disability.[45] To analyze anti-discrimination laws, it is important to understand what makes it different from other types of norms. According to T. Khaitan, head of research at the Bonavero Institute of Human Rights at the University of Oxford, for a norm to be considered "anti-discrimination" it has to satisfy four central conditions: it must require a connection between the act or the omission and a personal characteristic, it has to apply to grounds that can classify persons into one or more groups, there must be a situation in which members of one group (defined by the previous points) are significantly more likely to suffer disadvantage than members of at least another group, and lastly, the norm has to be designed in a way that it will benefit some (those who face discrimination) but not all members of the disadvantaged group.[46]

To formulate anti-discrimination laws and instruments, different subjects have adopted different approaches. In general, it is possible to observe three ways in which the grounds of such instruments can be determined.[47] First, we have the exhaustive list of grounds, found for example in the EU anti-discrimination legislation. This first approach relies on a list of grounds that are set and cannot be modified or extended by the judiciary system. However, this type of legislation allows for marginalized groups to pressure both the judiciary and the political levels to expand the protected grounds by claiming protection on the grounds included even if they do not fit their exact issue. For example, if the legislation includes sex but not sexual orientation, members of the LGBT community will claim that they are discriminated based on sex. This type of pressure, if powerful enough, can lead to a change in the legislation to include the "new" ground of discrimination.[48] This is what happened in the EU for example, where coverage of anti-discrimination legislation has been limited until very recently since the community was, for a long time, an economic agreement. Excluded groups tried to bring themselves within the grounds of discrimination laws, which led to mixed responses by courts but ultimately enlarged the grounds of the laws.[49] Politically, the pressure put on EU courts led to the fast adoption of two directives on the topic soon after the Treaty of Amsterdam gave competence on discrimination law to the EU, through which the principle of equal treatment

[45] Ibid. p. 111
[46] Khaitan, T. 2015, *A Theory of Discrimination Law*, Oxford, Oxford University Press, p. 42
[47] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 111
[48] Ibid. p. 113
[49] Ibid. pp. 114 – 115

was extended to prevent discrimination based on racial or ethnic origin, age, disability, religion, and sexual orientation.[50]

The second approach to anti-discrimination laws is the open-textured model. In this type of approach, opposite of the first one, the protection is based on an open-ended equality guarantee, such as the US Fourteenth Amendment: "*... No State [...] shall deny to any person within its jurisdiction the equal protection of the laws*".[51] This type of approach ensures that any type of provision can be challenged and it gives full power to the judiciary on the classification of illegitimate legislations and the determination of protected groups. In the case of the US, the epitome of this approach, the Supreme Court has full responsibility on these matters, and developed a specific approach to cases concerning discrimination, based on a double standard of scrutiny that include rationality and strict scrutiny in provisions that interfere with fundamental rights or operate to the disadvantage of specific groups.[52] Leaving the developing of grounds of discrimination to judges has, however, proven controversial and complex, especially in cases relating to race, gender, and sexual orientation to which strict scrutiny was not applied in early cases and discriminatory laws were considered legitimate under the rationality standard.[53]

The third and final approach is the non-exhaustive list approach, used for example in the European Charter of Human Rights and the Canadian Charter of Rights. This approach provides a list of possible grounds of discrimination but gives the judiciary system the possibility to extend the provided list if considered appropriate. This is achieved using expressions that indicate possible reasons for discrimination, for example the expression "such as" in the case of the ECHR or "including" used in the South African Constitution,[54] or specifying some of the grounds in a second part of the norm while providing a general criterion in the first part. This is the case for the Canadian Charter, in which we find the enunciation of the general equality principle, followed by the expression "*and in particular without discrimination based on race, national or ethnic origin, colour, religion sex, age or mental or physical disability*". [55] This approach facilitates the process of updating the list of possible grounds, as it is rare for a court to dismiss a case on the basis of not falling within the scope of the article, in turn, it also creates a situation in which grounds not specifically included are not regarded as equally severe, which can lead to the creation of hierarchies of scrutiny.[56] Although

---

[50] Council Directive 2000/43/EC of 29 June 2000 [2000] OJ L180/22
Council Directive 2000/78/EC of 27 November 2000 [2000] OJ L303/16

[51] U.S. Const. amend. XIV, Section 1

[52] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 118

[53] Ibid. pp. 119 – 121

[54] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) art 14

[55] *Canadian Charter of Rights and Freedoms,* s 15, Part 1 of the *Constitution Act*, 1982, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11

[56] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 126

anti-discrimination law has grown rapidly in the last decades, true equality is yet to be achieved. For some groups, enormous progress has been made, but at the same time it became clear that equal treatment before the law is not sufficient to eliminate inequalities, and that other types of guarantees are necessary to tackle indirect discrimination on workplaces, housing, access to services by prohibiting prejudiced behavior and discrimination from private actors, as well as providing accommodation or adjustment to ensure equal participation.[57]

1.1.2 A Competing Right: Freedom of Speech and its definition

Another central aspect of the debate over hate speech regulation relates to its apparent contrast with the right to freedom of speech. The right to freedom of expression is a fundamental human right that is protected in all human right systems. In general, there are different rationales behind the importance of freedom of speech. The right is considered essential to avoid abuses by states, as the suppression of free speech by the state is a clear and dangerous intrusion in one's individuality.[58] Moreover, the ability to express one's beliefs and ideas without fear of censorship or repercussions, is essential for the autonomy of the individual. Its suppression implies a lack of legitimate political debate, undermining democracy and pluralism, the absence of forms of expression that can help citizens form dissenting or critical ideas, like art and literature, as well as a threat to individual self-fulfillment deriving from one's ability to express autonomy.[59]

Freedom of speech is often connected to the role of promoting truth. In fact, the public debate on issues is necessary in order to pursue truth, for example when the principle is applied to journalism. However, the free speech principle could constitute a threat to the pursuit of truth if applied too broadly. In politics for example, the spread of false news regarding political figures can become a risk for the legitimacy of the political debate. Hence, the pursuit of truth does not only legitimize the protection of free speech, but also provides a justification and a reasoning for limitations on the right.[60] Many scholars argue against the possibility of freedom of speech being an absolute right. According to Abigail Levin for example, the argument that sees freedom of expression as maximizing public discourse is, in reality, a threat to such discourse. In this view, the lack of a regulation on free speech can potentially lead to the discrediting of the minorities in the society. In fact, if racist, sexist, homophobic speech is allowed and unsanctioned, the views and opinions of those minorities will

[57] Ibid. p. 154

[58] Fredman, S. 2018. "Freedom of speech" in *Comparative Human Rights Law,* pp. 305 - 354. Oxford: Oxford University Press, p. 307

[59] Ibid. p. 311

[60] Ibid. p. 309

have no importance nor resonance in the society, for they have already been discounted by the previous speech.[61]

Concerning the principle of truth, Justice Dikson provided an especially insightful opinion: the extremization of the principle of truth implies a complete lack of legislation on speech, as there is no absolute way to identify truth. However, according to Justice Dikson, the possibility that statements inciting hatred or discrimination are true is too little for considering them as crucial in the social and political debate and providing absolute protection for them.[62] Hence, the protection of free speech comes with its own challenges and difficulties. Therefore, to ensure the protection of the right, constitutional texts and international conventions apply different approaches that leave more or less space to interpretation. The oldest example, the US First Amendment, is the vaguest as it seems to grant an absolute right. The wording of the provision, stating only that the congress shall not make laws limiting freedom of speech and of the press, does not provide any limitation to the right, which results in a difficulty in balancing the right to freedom of expression against other rights, like protection from discrimination for example.[63] In more recent texts, like the International Covenant on Civil and Political Rights and the European Convention on Human Rights, the attributes and limits of the right are explicitly stated, leaving less space for interpretations.

In the European case, the second paragraph affirms that the right "carries duties and responsibilities" and provides a test to the balancing of such limitations, which must be prescribed by law and necessary in the democratic society.[64] This type of approach simplifies the interpretation of the provision, facilitating the balancing of rights against freedom of speech by the courts. Another approach, found in the Canadian Charter for example, provides a limitation clause that is general for the entire document and not specific to the right at question. In this case, the general clause provides that the rights in the charter are subject to "such reasonable limits prescribed by laws as can be demonstrably justified in a free democratic society". [65] What emerges from all approaches and iterations of the right is the need to balance it with other rights and the challenges that such balancing implies, which in turn makes the definition of standards and rules concerning hate speech especially challenging.

---

[61] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 65

[62] Brown, Alexander. 2015. *Hate Speech Law: a Philosophycal Examination.* New York: Routledge, p. 111

[63] Fredman, S. 2011, *Discrimination Law*, Oxford: Oxford university press, p. 306

[64] Ibid. p. 314

[65] *Canadian Charter of Rights and Freedoms,* s 1, Part 1 of the *Constitution Act*, 1982, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11

1.2 Hate speech in International Law

The issue of hate speech is a highly disputed matter among scholars, politician and in the public debate of the current time; however, in international law we can find instruments and provisions that allow or even require bans on certain types of speech, as a protection to the public order or for anti-discrimination purposes, seen freedom from discrimination is a fundamental idea of human rights. It is important to note that international law does not define "hate speech" directly in any treaty, but it has been defined by the United Nation in the Strategy and Plan of Action on Hate Speech.[66] Moreover, a distinction is made between three categories of Hate speech based on the severity of the speech act. The top-level hate speech acts are those already prohibited under international law. This category includes incitement to genocide and "*racial and religious hatred that constitutes incitement to discrimination, hostility or violence as defined in article 20 of the International Covenant on Civil and Political Rights*".[67] The second level, intermediate, includes certain forms of hate speech that may be prohibited under international law only if the limitations provided by law pursue a legitimate aim and are necessary and proportionate[68]. At the bottom level, the guidance places dissemination of lawful expressions that are, for example, offensive, shocking, or disturbing. These must not be subject to legal restrictions unless they also constitute incitement to discrimination, to avoid the shutting down of legitimate debate.

Nonetheless, the guidance also states that, although these speech acts do not fall into legal restrictions, states are encouraged to respond to them with different strategies and policies that should address the "*root causes, and counter the impacts of hate speech, including such forms at the bottom level.*".[69] These UN documents are not binding and therefore do not imply any obligation for States; however, The UN refers to other international documents that give some international law standards on hate speech, which will be analyzed in the following paragraphs.

1.2.1 Universal Declaration of Human Rights

The Universal Declaration of Human Rights was adopted by the United Nations General Assembly on 10 December 1948. The document was drafted by representatives of the fifty member states of the United Nations of the time reunited in the Commission of Human Rights with the guidance of Eleanor

---

[66] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech*, p. 8

[67] Ibid. p. 12

[68] Hogan Lovells, *The Global Regulation of Online Hate: A Survey of Applicable Laws – Special Report*, Hogan Lovells for PeaceTech Lab (December 2020), p. 8

[69] Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech, – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml, p. 15

Roosevelt. It is the first example of a declaration that seeks to universally protect fundamental human rights, and, over the years, it has been used as a reference in human rights matters. The declaration, although non-binding, is widely recognized in the international system and is cited in the preamble of over seventy human rights treaties. The declaration contains thirty articles that protect, along with other rights, the right to asylum, freedom from torture, free speech, education, and an array of civil, political, economic, social, and cultural rights.

In the Declaration there is no direct mention of hate speech, however, in the context of hate speech regulation, three articles of the Declaration are to be analyzed. The first article that has a clear connection to the matter is article 19: "*Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers*".[70] Article 19 does not provide for explicit prohibitions to the right to freedom of expression. However, two other articles provide limitations to the aforementioned right, the general limiting clause at Article 29, and Article 7 regarding discrimination. The latter expresses the principle of equality before the law and states that "*All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination*".[71] Therefore, it does apply to some hate speech acts, as it does entail that protection should be available for incitement to discrimination, and not only discrimination in itself. Moreover, Article 29 of the Declaration states that the exercise of the rights contained in the document may be subject to limitations determined by law for the purpose of "*securing due recognition and respect for rights and freedoms of others*".[72]

UDHR[73] drafters were aware of the risks of unregulated hateful and discriminatory speech, as the horrors of the Second World War had been uncovered and the connection between speech acts and the following events could not be ignored. Although the provisions alone do not expressly refer to speech acts, the history of these two provisions shows that the majority of the commission interpreted and understood the articles as implicitly allowing for restrictions limiting the right to freedom of speech in cases in which it amounted to incitement to hatred. In the first drafts of the Declaration, concerns regarding the potential damage resulting from the abuse of the right to free speech were expressed. The United Kingdom, for example, submitted a provision that allowed restrictions to the right to freedom of expression on publications that would have entailed a threat or an explicit attack to human rights and fundamental freedoms. The accompanying Comment, while acknowledging some concerns regarding the broadness of the provision, stated that the Bill of Rights would not have

[70] UN General Assembly. (1948). *Universal declaration of human rights* (217 [III] A). Paris. Art. 19 Available at: https://www.un.org/en/about-us/universal-declaration-of-human-rights
[71] Ibid. art 7
[72] Ibid. art 29
[73] Universal Declaration of Human Rights

prevented any government from taking steps against such publications, seen as the establishment of fundamental freedoms was the main goal of the Bill itself.[74]

Another example of the intentions of the committee can be found in the change that the original draft provision on freedom of expression underwent. The first draft contained the expression "*by any means whatsoever*" with regards to freedom of expression. However, after the representative of the Coordinating Committee of Jewish Organizations addressed the court, the expression was deleted as it was acknowledged that in no case freedom of expression should imply freedom of inciting hatred and discrimination.[75] When the draft reached the Commission on Human Rights, more concerns regarding the limitations to freedom of speech were expressed. It is important to notice how the Soviet representative proposed an amendment to the equal protection provision (Article 7) that explicitly prohibited advocacy for hatred. However, the proposal was firmly opposed by the U. S. through the chair of the commission, Eleanor Roosevelt, as the proposed amendment was considered impossible to apply. Moreover, concerns regarding the possible political implication of such a prohibition were expressed, seen as the Soviet proposal would have been applicable to political opinions as well. A debate ensued, but in the end the Soviet proposal failed.[76]

By understanding the position of the drafting commission, it is possible to state that the final wording of article 7 of the Declaration had the purpose of protecting against speech that would have amounted to national racial and religious hatred. It is also possible to say that, although there is no direct mention in Article 19 because of the debate in the Commission, freedom of expression was considered subject to Article 29, regarding abuses of rights.[77] Although the final Declaration did not contain explicit prohibitions, the drafters were aware of the risks and damages that propaganda and incitement to national, racial, and religious hatred could create, and this awareness was what pushed the Commission to draft the following binding documents that would have enabled the Declaration to be more than just words.[78]

1.2.2 International Covenant on Civil and Political Rights

The International Covenant on Civil and Political Rights was adopted by the UN General Assembly in 1966. In this convention, for which preparatory work started in the same commission that drafted the UDHR, Article 19 once again pertains to the right to freedom of expression. In this instance,

[74] Farrior, Stephanie. 1996. "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, pp. 1-98, p. 12-13

[75] Ibid. p. 14

[76] Ibid. p. 16

[77] Ibid. p. 19 – 20

[78] Ibid. p. 21

however, the article directly contains some restrictions for the aforementioned right and is therefore considered integral part of international law regarding hate speech. Paragraph 3 of the article states:

> The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
> (a) For respect of the rights or reputations of others;
> (b) For the protection of national security or of public order (ordre public), or of public health or morals.[79]

In the International Covenant it is expressly stated that the exercise of the right carries duties and responsibilities and therefore, the right is not absolute, it can be limited to protect the rights of others or to protect national security and public order.

Early drafts of this document prove how many of the drafters were especially concerned about the relation between advocacy of hatred and the phenomenon of discrimination, therefore considering incitement to discrimination and hatred as dangerous as incitement to violence in the context of the right to freedom of expression.[80] This consideration became an integral part of the convention in article 20(2): "*Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law*".[81] The discussion regarding this specific article was centered around the necessity to include incitement to hatred in the final version of the covenant as well as incitement to violence towards minorities. In general, two main lines of thought were expressed during the debate and related to the hierarchy of the two discussed matters, non-discrimination, and free speech. Part of the representatives, remembering the effects of Nazi public advocacy of hatred towards the Jewish community, believed that it was necessary to include incitement to hatred and hateful propaganda in the convention.[82] In opposition, other representatives focused on the abuses that terms such as "hatred" and "incitement" could bring, thus endangering the protection of free speech. As happened earlier for the Universal Declaration of Human Rights, during the discussions the Chair Eleanor Roosevelt expressed her concerns towards the extended version of the provisions limiting freedom of speech.[83] Once again, the intervention of a representative of the

---

[79] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, art. 19, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx
[80] Farrior, Stephanie. "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech", p. 26
[81] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, art. 20, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx
[82] Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, p. 28
[83] Ibid. p. 27

World Jewish Congress countered the arguments of many representatives, worried for the possibility of an abuse of the clause and for the difficulty in determining "incitement to hatred".[84] The Jewish representative clearly stated that it was necessary to include such a prohibition, seen as there were not enough guarantees in other documents or in other articles that hate propaganda would have been prohibited.

Finally, the commission adopted the Chilean amendment, which contained the "incitement to hatred" clause and the draft Covenant reached the Third Committee of the UN General Assembly.[85] Here, a similar discussion took place and more attempts to include further restrictions in article 19 regarding incitement to hatred and discrimination were made. In the end, the committee voted and approved a text proposed by sixteen countries which was considered a compromise between the two main lines of thought. In fact, the proposed text included the concept of incitement to hatred but, seen as the term alone was considered too broad, the committee agreed on the addition of the expression "*that constitutes incitement to discrimination, hostility or violence*" in order to pose a limit to what could be considered hatred.[86] The two articles of the ICCPR[87] that relate to hate speech, 19 (3) and 20 (2), which may seem in conflict with each other as one guarantees the right to freedom of speech while the other prohibits certain kinds of speech, have, in reality, been positioned one after the other with the specific intent of showing and emphasizing the relation between the two as a result of the debate in the Committee.[88]

To conclude the overview of the ICCPR, it is essential to analyze the approach of the Human Rights Committee, a body of independent experts created with the objective of monitoring the implementation of the covenant to which all state parties must submit regular reports. In general, the reports and General Comments issued by the Committee are considered as authoritative interpretations of the Covenant itself; therefore, the statements regarding articles 19 and 20 are useful tools in understanding of the meaning and the scope of the provision. In General Comment 11, the Committee has stated that the two articles are compatible since the prohibitions contained in article 20 (2) of the covenant are in line with the exercise of the right freedom of expression expressed in article 19, which carries by definition duties and responsibilities. In the comment the HR committee also underlines how any law that seeks to implement one of the two articles must be compatible with both and that "*For article 20 to become fully effective there ought to be a law making it clear that*

[84] Ibid. p. 28
[85] Ibid. p. 38 – 39
[86] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, art. 19, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx, art. 20
[87] International Covenant on Civil and Political Rights
[88] Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, p. 36

propaganda and advocacy as described therein are contrary to public policy and providing for an appropriate sanction in case of violation".[89]

The committee has repeatedly demonstrated through its General Comments that the obligations under the article are to be taken seriously. In General Comment 11 the committee effectively dismissed one of the main arguments against them, and that for the convention to be fully effective, States must create laws that clearly prohibit advocacy and sanction violations. Moreover, in carrying its reviews of individual country reports, the Committee urged different States to enact laws in order to make effectively prohibit national, religious and racial hatred.[90] Moreover, in General Comment No. 34 on article 19, the Committee provides a more detailed explanation on how the article should be interpreted in conjunction with article 20. According to the committee:

> What distinguishes the acts addressed in article 20 from other acts that may also be subject to limitations, is that for the acts addressed in article 20, the Covenant indicates the specific response required from the State: their prohibition by law.[91]

In fact, to ensure that freedom of speech is not unnecessarily restricted, the Committee states that:

> It is only with regard to the specific forms of expression indicated in article 20 that States parties are obliged to have legal prohibitions. In every case […] it is necessary to justify the prohibitions and their provisions in strict conformity with article 19.[92]

By this approach, it is possible to state that, for acts to be considered as falling in the scope of article 19 and 20 of the ICCPR in conjunction and considered punishable, they must amount to advocacy, therefore must be intended to elicit a response, they must exist for purposes of national, racial or religious hatred and they must constitute incitement to discrimination, hostility or violence.[93]

[89] UN Human Rights Committee (HRC), *CCPR General Comment No. 11: Article 20 Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred*, 29 July 1983, available at: https://www.ohchr.org/Documents/Issues/Opinion/CCPRGeneralCommentNo11.pdf

[90] Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, p. 45

[91] UN Human Rights Committee (HRC), *General comment no. 34, Article 19, Freedoms of opinion and expression*, 12 September 2011, CCPR/C/GC/34 , available at: https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

[92] Ibidem

[93] European Website on Integration: https://ec.europa.eu/migrant-integration/library-document/hate-crime-and-hate-speech-europe-comprehensive-analysis-international-law_en [Accessed 18 November 2021]

1.2.3 Convention for the Elimination of Racial Discrimination

The International Convention on the Elimination of All Forms of Racial Discrimination entered into force in January 1969 with the aim of abolishing racial discrimination. According to the convention, abolition should happen with two main strategies. In fact, the convention sets to first prohibit the phenomenon of racial hatred, and second, to prevent the issue as a whole by promoting education on the matter. As an International Law instrument concerning Hate Speech, this convention goes a step further compared to the previously cited Covenant, as the articles that most pertains to the issue at hand sets forward a mechanism that aims at preventing the issue of racial discrimination by partly using law as a deterrent.[94] In the convention article 4 and article 20 are the ones related to Hate Speech. As for previous conventions and declarations, the drafting of the CERD[95] posed some issues related to the apparent contrast between those who felt it would be necessary to prohibit the dissemination of racial superiority ideas and those who believed such prohibition would have been an unacceptable limitation of freedom of expression. As a result, article 4 of the convention was a compromise between the two positions:

> States Parties condemn all propaganda and all organizations which are based on ideas or theories of superiority of one race or group of persons of one colour or ethnic origin, or which attempt to justify or promote racial hatred and discrimination in any form, and undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of, such discrimination and, to this end, with due regard to the principles embodied in the Universal Declaration of Human Rights and the rights expressly set forth in article 5 of this Convention, inter alia:
>
> (a) Shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;
> (b) Shall declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination, and shall recognize participation in such organizations or activities as an offence punishable by law;

---

[94] *Committee on the Elimination of Racial Discrimination, General Recommendation VII relating to the implementation of Article 4 of the Convention, (Thirty-second session, 1985), Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies,U.N. Doc. HRI\GEN\1\Rev.1 at 64 (1994).* [online] Available at: http://hrlibrary.umn.edu/gencomm/genrevii.htm
[95] International Convention on the Elimination of All Forms of Racial Discrimination

(c) Shall not permit public authorities or public institutions, national or local, to promote or incite racial discrimination. [96]

The original draft provided for legal punishments for all incitement acts that would have resulted in or would have likely caused acts of violence.[97] However, when the convention reached the UN general assembly, the wording of the article seemed too broad. The UK, one of the opposing states, announced that it "*would not accept punishment for ideas or incitement unless there was incitement to violence*".[98] The final compromised text was proposed by Nigeria, and stated "*all acts of violence or incitement to such acts by any race or group of persons*". Moreover, the Nigerian representatives proposed the so called "due regard" clause, by virtue of which parties agreed to adopt measures to eradicate incitement or violent acts with due regard to the principles of the universal declaration. By inserting this clause in the final text, the universal declaration becomes a balancing tool, for all rights stated in the declaration are to be given "due regard", thereby allowing the protection freedom of expression and freedom of association. However, the same clause has been interpreted by some states, namely the United States or the United Kingdom, as not imposing any obligation on states to act against actions in a way that could interfere with the above-mentioned rights.[99]

The conventions also institute a body of independent experts tasked with the monitoring of the implementation of the convention itself: the Committee on the Elimination of Racial Discrimination. The committee examines the reports submitted by states parties every two years, addresses concerns and gives recommendations. Moreover, the committee can intervene as a monitoring body by analyzing interstate complains and individual complains, or by engaging in "*preventive measures which include early-warning aimed at preventing existing situations escalating into conflicts and urgent procedures to respond to problems requiring immediate attention to prevent or limit the scale or number of serious violations of the Convention*"[100]. Finally, the committee periodically publishes its comments and interpretations concerning human rights provisions.[101]

One of the reports by the committee, published in 1986, refers specifically to the due regard clause and specifies that the clause is not to be interpreted as a cancellation or a departure from the

[96]UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, p. 195, available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx

[97] Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, p. 48

[98] Ibid. p. 49

[99] Ibid. p. 50

[100] United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/HRBodies/CERD/Pages/EarlyWarningProcedure.aspx [Accessed 22 November 2021]

[101] United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/HRBodies/CERD/Pages/CERDIntro.aspx [Accessed 22 November 2021]

obligations set forth in paragraphs a and b of article 4 seen as, if that was the case, the obligation would have had no reason of being included in the convention. Moreover, the committee emphasized that freedoms are not absolute, as the existence of society itself implies that freedom of actions by members is limited.[102] Such interpretation was reinforced by referring to article 29 and 30 of the UDHR, and to article 19 of the Covenant. By providing this interpretation the committee drifted away from the more liberal interpretation given by the US and the UK, among others, thus showing a propensity to give greater weight to the right to freedom from racial discrimination compared to the right to freedom of expression.[103]

More recently, in 2013, the Committee issued General recommendation No. 35, on combating racist hate speech, emphasizing that member states should honor their obligations under CERD.[104] In the recommendation the committee reiterated the fact that drafters of the convention were aware of the contribution of speech in creating a favorable climate for hatred and discrimination and, even though the term "hate speech" is not contained in the convention, the committee focused on the provisions that "*enable the identification of expression that constitutes hate speech*".[105] In the recommendation, the mandatory nature of article 4 and the expression "*State parties [...] undertake to adopt immediate and positive measures*"[106] are underlined, as well as the function of prevention and deterrence that are inherent to the article. By this interpretation, article 4 is understood as effectively criminalizing forms of racist hate speech that falls within its scope, namely:

a) All dissemination of ideas based on racial or ethnic superiority or hatred, by whatever means;
b) Incitement to hatred, contempt or discrimination against members of a group on grounds of their race, colour, descent, or national or ethnic origin;
c) Threats or incitement to violence against persons or groups on the grounds in (b) above;
d) Expression of insults, ridicule or slander of persons or groups or justification of hatred, contempt or discrimination on the grounds in (b) above, when it clearly amounts to incitement to hatred or discrimination;

[102] Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, p. 50
[103] Ibid. p. 52
[104] Pálmadóttir, Jóna Aðalheiður. Kalenikova, Iuliana. 2018. "*Hate speech: an overview and reccommendations for combating i*t", Reykjavik: Icelandic Human Rights Centre, p. 9
[105] UN Committee on the Elimination of Racial Discrimination (CERD), *General recommendation No. 35: Combating racist hate speech*, 26 September 2013, CERD/C/GC/35, para. 5
[106] UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, p. 195, available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx

e) Participation in organizations and activities which promote and incite racial discrimination.[107]

However, since the article is not self-executing, the committee affirms that States parties are to adopt legislation to combat the previously mentioned sanctionable types of hate speech. Moreover, the committee underlined some contextual factors that should be considered for the qualification of punishable conducts. These factors will be later reiterated by the Rabat Plan of Action, as for example the content and form of speech, the context in which the speech was delivered, the status of the speaker, the reach, and the objectives of the speech. All these characteristics should be analyzed, seen as, for example, a discourse that could be innocuous in one context, could become much more significant in another or if delivered by a politician.

Regarding the due regard clause, the recommendation reiterates the previously published recommendation and affirms that "*the relationship between proscription of racist hate speech and the flourishing of freedom of expression should be seen as complementary and not as a zero-sum game*"[108] and that the priority given to one of the two does not imply a diminution of the importance or relevance of the other.

1.2.4 Relevant UN documents: the Rabat plan of Action and the reports of the Special Rapporteur on Freedom of Opinion

The last part of this international law analysis focuses on a few relevant UN documents that contribute to the regulation of Hate Speech phenomena in the international system. Firstly, the analysis will center on the Rabat Plan of Action on the prohibition of advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. This plan of action is the result of a series of expert workshops organized by the Office of the High Commissioner for Human Rights in 2011 and 2012, which analyzed and explored legislative and judicial patterns, as well as policies that related to the prohibition of incitement to hatred based on nationality, *race,* or religion. It has gained some attention within and outside of the UN after its publication, as it is considered by many as a highly significant tool in the understanding and implementation of international law on freedom of expression and incitement to hatred.[109] At the same time, the plan has been scarcely

[107] UN Committee on the Elimination of Racial Discrimination (CERD), *General recommendation No. 35: Combating racist hate speech*, 26 September 2013, CERD/C/GC/35, para. 13
[108] Ibid. para. 45
[109] Molnar, P. Ed. 2015, *Free Speech and Censorship Around the Globe*, Budapest, Central European University Press, p. 213

acknowledged by those that should be the main actors and subjects of the recommendations, mostly states.

The Rabat plan finds its origins in the continuously evolving UN approach to antiracism, as well as in the longstanding debate that has been carried in the UN human rights bodies regarding freedom of expression and hateful speech. The outcome document of the Durban Review Conference of 2001 was the first document that acknowledged the will to organize such expert workshops about incitement to hatred and its characteristics in different countries and regions.[110] Four regional workshops were organized and comprehended 55 experts from different fields of expertise, especially in relevant areas of law and policy, such as free speech and equality. The workshops were organized as follows: a presentation of regional study and trends in legislation regarding incitement, joint submissions by the Special Rapporteur on freedom of opinion and expression, expert papers by members of the treaty bodies, academical papers, NGO representatives' papers. The workshops proceeded independently, and the findings were analyzed and summarized at the final meeting in Rabat. At said meeting, the experts also proceeded to identify possible next steps in the prohibition of advocacy to hatred.[111]

The Rabat Plan of Action characterized itself as a document strongly rooted in international law, thanks to the central role of UN treaty bodies and international NGOs played in the workshops. Moreover, the RPA[112] has a strong focus on the problems that relate to responses and remedies regarding incitement to hatred, which makes it a potentially central tool in future approaches to the issue. The Plan of Action is divided into three different sections: legislation, jurisprudence, and policies. For each section the document provides conclusions and recommendations. In the first section, legislation, the document underlines how anti-incitement laws, where present, "*are frequently "heterogeneous, at times excessively narrow or vague*," reliant on "*variable terminology*" which is often at odds with Article 20 of the ICCPR or excessively broad, thus "*[opening] the door for arbitrary application of the laws*".[113] The conclusions highlight how, in international law, the prohibition of incitement to hatred is clearly established by article 20 of the ICCPR, which requires a high threshold to ensure that limitations of speech remain an exception, but also implies the need for specific and proportionate restrictions.[114] Moreover, the document highlights how, in general,

---

[110] United Nations website: https://www.un.org/WCAR/e-kit/backgrounder1.htm [Accessed 15 November 2021].

[111] Molnar, P. Ed. 2015, *Free Speech and Censorship Around the Globe*, Budapest, Central European University Press, p. 220

[112] Rabat Plan of Action

[113] Molnar, P. Ed. 2015, *Free Speech and Censorship Around the Globe*, Budapest, Central European University Press, p. 223

[114] UN Human Rights Council, *Annual report of the United Nations High Commissioner for Human Rights : Addendum, 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*, 11 January 2013, A/HRC/22/17/Add.4, para. 17 – 18, available at: https://www.ohchr.org/en/issues/freedomopinion/articles19-20/pages/index.aspx

anti-incitement laws seem to be used to prosecute minorities instead of real incitement cases, with very low rates of judicial mechanisms being used in such cases.[115]

Concerning the legislation, the plan of action recommends, among other things, the necessity for a clear distinction between:

> […] expression that constitutes a criminal offence, expression that is not criminally punishable but may justify a civil suit or administrative sanctions; expression that does not give rise to criminal, civil or administrative sanctions, but still raises concern in terms of tolerance, civility and respect for the rights of others.[116]

Moreover, states are encouraged to ensure that the three-part-test applies to cases of incitement to hatred, and to "*adopt comprehensive anti-discrimination legislation that includes preventive and punitive action to effectively combat incitement to hatred*".[117] For what concerns the second section, the conclusions state that often there is no recourse to judicial mechanisms, that victims are mostly from disadvantaged groups and that case law on incitement to hatred is not easily available. They blame the lack of adequate legislation and judicial assistance.[118] To combat these issues the recommendations, the RPA suggests that national and regional courts should be regularly updated about international standards and jurisprudence. Relating to individuals, the document highlights how states should ensure effective remedy through civil or non-judicial remedy, as well as through administrative sanctions, civil sanctions, and criminal sanctions as a last resort measure.[119] Lastly, regarding the third section on policies, the RPA gives different recommendations to different actors involved in the issue. For states, the focus is on combating stereotype and promoting intercultural understanding, while creating an adequate legal infrastructure and ensuring that the regulatory framework promotes pluralism and non-discrimination in media. Finally, strengthening the current international human rights mechanisms and UN treaty bodies should be a priority for states.[120]

Although possibly being an extremely useful tool, the RPA has some critical points. First, the plan of action does not have a comprehensive approach to the issue of incitement to hatred, since it is heavily reliant on article 20 of the ICCPR and fails to provide other grounds of discrimination, leaving out sex or sexual orientation. Second, the Plan of action does not consider the divergence between article 20 of the ICCPR and article 4 of the CERD therefore missing the opportunity for providing a clear

[115] Ibid. para. 11
[116] Ibid. para. 20
[117] Ibid. para. 26
[118] Ibid. para. 28
[119] Ibid. para. 33 – 34
[120] Ibid. para. 51 – 53

interpretation for the application of those two conventions. Online hate speech is not mentioned in the document, which is a major issue considering its growing importance and the undoubted difficulties states have in developing appropriate responses to the issue.[121]

To conclude the overview on relevant UN documents relating to hate speech, it is important to mention two reports of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, a mandate established in 1993 by the United Nation Commission on Human Rights (now Human rights council) with the purpose of protecting and promoting freedom of opinion and expression in light of international human rights law and standards.[122] The report dated 7 September 2012, titled "Promotion and protection of the right to freedom of opinion and expression", focuses on freedom of opinion and incitement to hatred and provides, along with an overview of existing international norms and standards for the right in question, a distinction between different types of hate speech and a threshold to assess the seriousness of the offence. In reporting the international norms and standard, the rapporteur states that "*International human rights law [...] recognizes that the right to freedom of expression can indeed be restricted when it presents a serious danger for others and for their enjoyment of human rights*",[123] thus recognizing the legitimacy, or even the necessity, of hate speech regulation. The report then underlines how the language of article 20(2) of the ICCPR explicitly prohibits any advocacy for national, racial, or religious hatred as does the Convention on the Elimination of All forms of Racial Discrimination. In giving this overview the report also highlights how restrictions imposed on the right to freedom of expression must comply with the three-part test to limitations of the Covenant: the restriction must be provided by law, proven necessary and legitimate by the state, and proven to be the least restrictive and proportionate mean to achieve the aim.[124]

After the overview of the international standards, the rapporteur focuses on clarifying the terms used by the covenant in order to avoid misapplication of the law and states "*first, only advocacy of hatred is covered; second, hatred must amount to advocacy which constitutes incitement, rather than incitement alone; and third, such incitement must lead to one of the listed results, namely discrimination, hostility or violence*". As a result, "*advocacy of hatred on the basis of national, racial or religious grounds is not an offence in itself*".[125] In the final part of the overview, the report provides

[121] Molnar, P. Ed. 2015, *Free Speech and Censorship Around the Globe*, Budapest, Central European University Press, p. 214

[122] United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/en/issues/freedomopinion/pages/opinionindex.aspx [Accessed 2 December 2021].

[123] UNGA, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/67/357 (7 September 2012), para. 37

[124] Ibid. para. 41

[125] Ibid. para. 43

a threshold, elaborated first by the NGO Article 19, which proposes a seven-part test for expression that follows these steps:

- Severity of the hatred, of what is said, the magnitude, the intensity, the media reach.
- Intent of the speaker to incite
- Content or form of the speech, degree to which it is provocative or direct.
- Extent of the speech in terms of reach
- Probability of harm occurring.
- Imminence of the acts called for by the speech
- Context in which the speech is delivered.

The special rapporteur then provides a list of essential elements in determining whether an expression can be considered incitement to hatred. These elements can be summarized as: real and imminent danger of violence, intent of the speaker to incite discrimination, consideration of the context in which the hatred was expressed (given that international law prohibits some forms of speech for the consequences and not necessarily for the content, seen as some expression might be extremely offensive in one culture but not in another). Relating to the context, attention must be put on the existence of patterns of tension between groups, the tone of the speech and the person and means of those inciting hatred.[126] Finally, the report highlights that while states are required to prohibit advocacy of hatred, there is no requirement for a criminalization of such expression, and in any case hate speech laws should never penalize speech that conveys true statements, that is dissemination of hate but has no intention of incitement to discrimination or violence, that is provided by journalists who have the right to decide how to communicate information and ideas to the public. Moreover, no one should be subject to prior censorship and in any case the imposition of sanctions must be in conformity with the principle of proportionality.[127]

In a more recent report by the Special Rapporteur, dated 9 October 2019, the focus shifts towards state action and content moderation by companies, and an entire section of the report is dedicated to online hate speech. It is interesting to note how, in this report, the overview of relevant international norms and standards is titled “Hate Speech regulation in international human rights law”, whereas in the previously cited report the section’s title referred to “incitement to hatred”.[128] In a way, the denomination of the section implies the recognition of a kind of legal value to the term hate speech, although the two words are still in quotes. The report reiterates the overview of international law

---

[126] Ibid. para. 46

[127] Ibid. para. 50

[128] UNGA, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/74/486 (9 October 2019), para. 4

norms and standard, also referring to the previously cited RPA and the factors identified in it for the determination of a threshold of severity. In the report, the rapporteur recommends that states apply human rights law for online hate speech as they would for offline speech, define the terms that constitute prohibited speech and content under international law documents, adopt laws that require companies to define standards and enforce rules against hate speech, and establish judicial mechanisms that ensure access to justice and remedies for individuals suffering harm from hate speech acts.[129] At the same time, the report once again reiterates how the criminalization of such acts is not the only option; hence, states should employ government actions to prevent and reduce hate speech acts, actively reducing the need for explicit bans on expression or direct sanctions on individuals.[130]

The report is especially hard in its opening statement of the recommendations for companies section, stating that "*Companies have for too long avoided human rights law as a guide to their rules and rule-making, notwithstanding the extensive impacts they have on the human rights of their users and the public*".[131] The rapporteur lastly gives a set of recommendations, stating that companies should evaluate how the products and services they offer affect the human rights of users and the public, adopt policies and rules on hate speech directly tied to international human rights law and standards, define the categories they identify as hate speech with reasoned and publicly provided explanations, ensure that hate speech rules are enforced through an evaluation of the context and harm of the speech involving the harmed community, and lastly, develop tools that promote free expression while at the same time ensuring de-amplification, de-monetization, education and counter-speech in their platforms.[132]

1.3 Regional Hate Speech Law and Approaches

To further deepen the understanding of international laws relating to the issue of hate speech, it is necessary to analyze the regional frameworks that have been put in place in different areas of the world. This kind of regional laws and recommendations can have a direct influence on single states and on the effective enjoyment of the right or the penalization of certain forms of expression. In this paragraph the analysis will focus on two regional frameworks, the European and the American, seen as these are the two regions where a hate speech framework has been developed by the regional organisms of the area. The first paragraph will center on the European situation, with the two main

---

[129] Ibid. para. 57
[130] Ibidem
[131] Ibid. para. 58
[132] Ibidem

organization, the council of Europe and the European Union. The second paragraph will focus on the Interamerican commission on human rights and its declarations and reports, non-binding instruments that can, however, help orienting the states of the region and the following decisions by courts or other related organisms.

1.3.1 European Hate Speech Law: Council of Europe and the European Union

The European Hate Speech framework of laws and relevant documents is, now, the most elaborate and complete. The first important contribution comes from the Council of Europe: recommendation No. (97) 20 of the Committee of Ministers to Member States dated 30 October 1997. In this recommendation that explicitly refers to hate speech, the committee first provides a useful definition of the term and of the scope of the recommendation. Hate Speech is defined as:

> […] all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerant expression by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.[133]

The report than goes to outline seven principles that the governments of member states and public authorities in general should follow in relation to hate speech, especially that delivered through the media. According to these principles, states have the responsibility to refrain from statements that could be reasonably understood as hate speech or as legitimizing hatred.[134] At the same time, states should establish a strong legal framework for hate speech cases and ensure that it does not interfere with the right to freedom of expression other than in a circumscribed and lawful manner.[135] Lastly, it is notable how principle 6 and 7 underline how national law and practice should carefully consider the role of media and distinguish between the author of the hate speech act and the responsibility of media professional to the dissemination of it as part of their mission to spread information, seen as reporting on racist, anti-Semitic, xenophobic and in general discriminatory acts is fully protected under article 10 of the European convention of human rights.[136]

---

[133] Council of Europe Committee of Ministers, Recommandation No. R (97) 21 of the Committee of Ministers to member states on "Hate Speech" (Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies), Scope
[134] Ibid. principle 1
[135] Ibid. principles 2 – 3
[136] Ibid. principles 6 – 7

Recommendation (97) 21 by the same body further enlarges the Council of Europe Approach to the matter. This particular document, "on the Media and the Promotion of a Culture of Tolerance", adopted on the same day highlights how the involvement of media in the issue of intolerance can make a notable positive contribution and foster understanding and respect on a large scale.[137] The last non-binding document is the Declaration of the Committee of Minister on freedom of political debate in the media, adopted in 2004, in which it is emphasized that freedom of political debate is not absolute and does not include opinions that are a clear incitement to hatred, racism or in general intolerance.

Another central development in the Council of Europe framework is the Additional Protocol to the Convention on Cybercrime, adopted on 28 January 2003. The protocol defines racist and xenophobic material as any written material, image or representation of ideas that advocates, promotes, or incites hatred and states that states parties are bound to criminalize "*distributing, or otherwise making available racist and xenophobic material to the public through a computer system*"[138]; threatening and commission of a criminal offense or insulting persons for the reason that they belong to a group defined by race, color, religion, descent or national origins; aiding the commission of such acts. The criminalization should, according to this document, also apply to the distribution of material which denies, justifies, minimizes, or approves of acts of genocide or crimes against humanity. A prerequisite for the criminalization in any of the aforementioned situations is, however, the intent of the act: parties are required to adopt legislative measures to establish such types of offences when those are committed intentionally and without right. Two other conventions worth mentioning in the framework of the Council of Europe are the revised European Social Charter, in which any discrimination on grounds of race, color, religion or nationality is prohibited, and the Framework Convention for the protection of national minorities, whose state parties are required to adopt adequate measures to effective equality and to encourage tolerance and dialogue to promote respect.[139]

Other than the committee of ministers, also the parliamentary assembly of the council adopted some documents related to the issue: resolution 1510(2006) in which it is once again highlighted that freedom of speech and hate speech against groups are not compatible; and Recommendation 1805 (2007) that recalls the necessity to penalize statements that clearly call for a hateful act or incite such acts. In addition, the council established the European Commission against Racism and Intolerance with the mission of combatting racism and discrimination through general policy recommendations

---

[137] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 10
[138] European Website on Integration: https://ec.europa.eu/migrant-integration/library-document/hate-crime-and-hate-speech-europe-comprehensive-analysis-international-law_en [Accessed 18 November 2021]
[139] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 7

and guidelines to member states. ECRI[140] released general policy recommendation n. 15 in 2015, On Combating Hate Speech, in which a definition of hate speech was provided:

> […] the use of one or more particular forms of expression – namely, the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or group of persons, as well any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression – that is based on a non-exhaustive list of personal characteristics or status that includes "race", colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation.[141]

Additionally, different recommendations regarding the application or creation of hate speech laws, the necessity to ratify treaties and the awareness raising approach states should have to the issue were made.[142] To have a complete understanding of the regional framework however it is necessary to analyze the European Union law on hate speech. The EU is one of the most active actors in addressing hate speech and hate crime in its legislation and overall organization. The first efforts to reduce racism and xenophobia date back to the early 1990s, as the awareness of the challenges that such behaviors pose to in the society grew. The first acts adopted focused primarily on racism, xenophobia and anti-Semitism and were adopted by the European Parliament and the Council in the form of a resolution ("racism, xenophobia and anti-Semitism" 1995)[143] and a Joint Action (1996)[144] in which the two bodies required and encouraged member states to act and ensure the availability of laws and measures to combat this kind of situations. These measures outlined the EU approach to the issue and expressly referred to "*public incitement to discrimination, violence or racial hatred in respect of a group of persons or a member of such a group defined by reference to colour, race, religion or national or ethnic origin*"[145] and "*public condoning, for a racist or xenophobic purpose, of crimes against humanity and human rights violations*"[146] as the first two behaviors that member states needed to ensure were punishable in order to cooperate on the matter.

---

[140] European Commission against Racism and Intolerance
[141] Council of Europe, European Commission against Racism and Intolerance (ECRI) General Policy Recommendation No. 15 on combating hate speech adopted on 8 December 2015, p. 16
[142] Ibid. p. 32 – 40
[143] European Parliament resolution, B4-0731/95 of 27 April 1995, Resolution on racism, xenophobia and anti-Semitism, Official Journal C 126, 22/05/1995 P. 0075
[144] Council of the European Union, Joint action 96/443/JHA, concerning action to combat racism and xenophobia, OJ L 185, 24.7.96, *EU Anti-Discrimination Policy: Annexes,* Available at: https://www.europarl.europa.eu/workingpapers/libe/102/text5_en.htm#annex4
[145] Ibid. para. A
[146] Ibidem

Later, in 2008, the Council adopted Framework Decision 2008/913/JHA on combatting certain forms and expression of racism and xenophobia by means of criminal law and repealed the previous Joint Action. The provisions contained in the CFD, to which member states were obliged to comply, provided for the creation of laws and regulations on racist and xenophobic hate speech and for the "*necessity to consider racist and xenophobic motivation as an aggravating circumstance or to ensure that courts take such motivations into account in the determination of penalties*".[147] Other than obliging member states to take the necessary measures to ensure crimes of incitement to violence or hatred are punishable by law, the CFD also includes the necessity to adopt laws against the condoning, denying, and trivializing of crimes against humanity and war crimes. As is expected for a binding document, in the case of the Framework Decision the threshold for speech to be considered hate speech is significantly higher compared to the UN level instruments and, seen as the decision does not account for civil tools to tackle hate speech, a lower threshold is impossible to have, as the consequences for hate speech are necessarily severe.

Moreover, a major flaw of the decision is the difficulty in uniformly applying it throughout the EU, due to the different national approaches to the definition of Hate Speech, to the limitations applicable to free speech and the lack of a European definition of the concept. Lastly, to be punishable, the speech must be racist or xenophobic, which leaves out homophobia, transphobia, and sexism from the decision and, consequently, all the hateful manifestations that could come from such types of discrimination.[148] The European Parliament has later recognized some of the issues in the decision, and issued the EU Roadmap against homophobia and discrimination on grounds of sexual orientation and gender identities, in which it was suggested that:

> The Commission should monitor and provide assistance to the Member States with regard to issues specific to sexual orientation, gender identity and gender expression when implementing Directive 2012/29/EU establishing minimum standards on the rights, support and protection of victims of crime.[149]

and that:

---

[147] Eur-lex website: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l33178 [Accessed 4 December 2021]

[148] Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, pp. 203 – 229

[149] European Parliament website: https://www.europarl.europa.eu/doceo/document/TA-7-2014-0062_EN.html [Accessed 3 December 2021]

> [it] should propose a recast of the Council Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law including other forms of bias crime and incitement to hatred, including on grounds of sexual orientation and gender identity.[150]

More recently, the Parliament reiterated the commitment against hate speech and especially homophobic hate speech, and especially homophobic and transphobic hate speech, through the resolution of 18 December 2019 on public discrimination and hate speech against LGBTI people, including LGBTI free zones. In the resolution expresses a deep concern regarding the discrimination against LGBTI people and calls on the European Commission and Member States to use all the tools and the procedures at their disposal to diminish discrimination towards the community and to enhance the legal protection to anyone on all grounds. Regarding hate speech, the recommendation calls on member states to monitor hate speech, with a special focus on public authorities and officials, and take concrete measures and sanctions against it.[151]

The last tool in the European legal framework is the EU Code of conduct on countering illegal hate speech online, negotiated by the commission with major online companies such as Facebook, Microsoft, Twitter and YouTube in 2016 and later joined by Instagram, Snapchat and Tik Tok. The code of conduct has the objective of countering the spread of illegal online hate speech and to help users identify and notify hate speech on social media, to tackle the phenomenon more efficiently. The central provision of the code of conduct is the guarantee of the review of the majority of valid notifications and the removal of the content in question, if deemed necessary, within 24 hours.[152] The last evaluation of the code of conduct found that, overall, the agreement had positive result and companies did assess most of the content within 24 hours and remove the "illegal" content. There was, however, a decrease in the average of notifications reviewed: 81% in 2021, compared to the 90,4% of 2020. The data in the report also allowed to highlight how sexual orientation and xenophobia are the most reported kinds of hate speech, respectively 18,2% and 18%.[153] The code of conduct requires companies to define rules and standards for hateful content and incitement internally, therefore, the definitions are different depending on the platform user's access. Although this might be seen as an improvement, seen as most companies included gender, sexual orientation, disability, in their definition, the fact that the rules and standards are provided by the company allows them to

---

[150] Ibidem.

[151] European Parliament resolution, 2019/2933(RSP) of 18 December 2019 on public discrimination and hate speech against LGBTI people, including LGBTI free zones, P9_TA(2019)0101, p. 8

[152] European Commission website: https://ec.europa.eu/newsroom/just/items/31811/en [Accessed 3 December 2021]

[153] European Commission, Factsheet – 6th evaluation of the Code of Conduct, Directorate-General for Justice and Consumers, 7 October 2021, pp. 1 – 3

verify content privately and, with the absence of clear standards, there is little space for appeal or review of the decisions.[154]

1.3.2 Interamerican Commission on Human Rights Hate Speech Law

In the Interamerican human rights system, as opposed to the European system, there are few provisions and instruments related to hate speech, possibly because the central actor of the area (the United States of America) has a well-established and broad interpretation of the right to free speech as being almost absolute.[155] Although the presence of such a strong actor with such clear views on the matter implies the lack of a complete and clear regional framework, there are some instruments that justify and acknowledge the existence of speech that does not fall under the protection of the right to free speech. The first regional instrument in this context is the American Convention on Human Rights, in which article 13 refers to freedom of expression and its limits. In fact, paragraph 2 and 5 of the article state respectively:

> The exercise of the right provided for in the foregoing paragraph shall not be subject to prior censorship but shall be subject to subsequent imposition of liability, which shall be expressly established by law to the extent necessary to ensure: a. respect for the rights or reputations of others; or b. the protection of national security, public order, or public health or morals.[156]

And:

> Any propaganda for war and any advocacy of national, racial, or religious hatred that constitute incitements to lawless violence or to any other similar action against any person or group of persons on any grounds including those of race, color, religion, language, or national origin shall be considered offenses punishable by law.[157]

The formula used in the paragraph is very similar to that used in previously analyzed conventions and instruments and requires states parties to effectively outlaw these kinds of expression. Moreover, the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights also provided declarations regarding the issue of Hate Speech, namely the joint statement, along with

---

[154] Casarosa, F. *Handbook on Techniques of Judicial Interaction in the Application of the EU Charter – Freedom of Expression and Countering Hate Speech,* European University Institute, Robert Shuman Centre
[155] Carlson, C. R. 2021, *Hate Speech*, Cambridge, The MIT Press, p. 126
[156] Organization of American States (OAS), *American Convention on Human Rights, "Pact of San Jose", Costa Rica*, 22 November 1969, available at: https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm
[157] Ibidem

the United Nations Special Rapporteur and the OCSE representative on freedom of the Media. In the statement they recognized that expression inciting violence and hate is harmful and it usually precedes or accompanies crimes against humanity. For this reason, according to the statement, it should be, when legitimate and necessary, regulated by laws.[158]
The Interamerican Commission on Human Rights, an autonomous organ of the Organization of American States, the main regional organization of the area, was created in 1959 and it is a central part of the Inter-American human rights system. The Inter-American Commission function is, in fact, promoting the observance and defense of human rights in the Americas and it is the primary organ in the process of analyzing individual petitions and alleged violations of human rights in the region.[159]
The commission created the office of the Special Rapporteur for freedom of expression in October 1997, which issues annual reports and included a section on hate speech in its 2015 report. In chapter IV of the report, "Hate speech and incitement to violence against lesbian, gay, bisexual, trans and intersex persons in the Americas", the office provides an overview of the inter-American legal framework on hate speech and incitement and an interpretation of the provisions of the American Convention on Human Rights on the matter.[160] The report underlines how in principle, all forms of speech are protected by freedom of expression and not only the ideas that are received favorably by the public. However, the right is not absolute and can be limited in cases of necessity.
The report clarifies the distinction between two paragraphs of article 13 of the Interamerican Convention. In paragraph 13(2), which concerns intolerant expression or comments, freedom of expression could be subject to subsequent liability to ensure that the rights of other individuals or groups are protected. Conversely, under paragraph 13(5), states *are required* to adopt laws that punish advocacy of hatred. The report proceeds by stating that, in the view of the special rapporteur and the IACHR, states should take action to guarantee enjoyment of freedom of expression by LGBTI persons, as a countermeasure for hate speech and an indirect deterrence for the phenomenon. However, according to the report, for "*Negative or derogatory portrayal and other expressions that stigmatize LGBTI persons*" "*the legal prohibition [...] will not do away with the stigma, prejudice, and hatred against LGBTI persons that is deeply rooted in the societies of the Americas.*" And "*more should be done to promote a comprehensive approach that goes beyond legal measures and includes preventive and educational mechanisms and measures implemented by States, media, and society in*

[158] Organization of American States website:
https://www.oas.org/en/iachr/expression/showarticle.asp?artID=443&lID=1https://www.oas.org/en/iachr/expression/showarticle.asp?artID=443&lID=1 [Accessed 5 December 2021]
[159] Organization of American States: https://www.oas.org/en/iachr/mandate/Basics/intro.asp> [Accessed 5 December 2021]
[160] IACHR, *Annual Report of the office of the Special Rapporteur for Freedom of Expression*, Office of the Special Rapporteur for Freedom of Expression, OAS. Official records; OEA/Ser.L/V/II, (2015)

general.”.[161] For an expression to be punishable, according to the special rapporteur, the person must have a clear intention of promoting violence, the capacity to achieve it and to create a real risk of harm. The report then reiterates that any limit to freedom of expression must meet the guarantees of being applied by an independent executive body, must respect due process, and must be accompanied by proportionate sanctions.[162]

The other instrument that needs to be mentioned is the Inter-American Convention against all forms of discrimination and intolerance, adopted on 6 June 2013 but entered into force in February 2020 after the second ratification was deposited. The convention was mandated by the OAS general assembly in 2005, and the negotiation process should have resulted in a single document, the Convention against racism and all forms of discrimination and intolerance. The draft convention included many different grounds of discrimination, such as sexual orientation, gender identity, educational level, migrant status, that are not explicitly recognized in key international human rights law documents. The negotiations lead to a deadlock that could be solved only by dividing the text. The outcome therefore comprised two different conventions, the anti-racism, and the anti-discrimination conventions. The division of the two texts implies that state parties can decide to ratify one (the anti-racism convention with all probability) and not the other, limiting the scope of the anti-discrimination provisions. Nonetheless, the anti-discrimination convention proves to be an useful instrument in combatting discrimination and intolerance, as well as incitement. In fact, article 4 of the convention affirms that state parties undertake to prevent, prohibit, and punish publication or dissemination in any form of materials that advocate or incite for hatred, or condone or justify acts of genocide or crimes against humanity.[163]

1.4 European Court of Human Rights Jurisprudence on Hate Speech

To discuss the jurisprudence of the European Court of Human Rights in cases of hate speech an overview of the applicable provisions of the European Convention on Human Rights is necessary, as it is the most important regional human rights instrument, and it is the convention itself that established the court and its jurisdiction. The two relevant articles, (and the ones that are mostly appealed to in the applications to the court), are article 10 and 17. Article 10 is the one relating to freedom of expression. Paragraph 1 of the article states that freedom of expression “*shall include*

---

[161] Ibid. pp. 366 – 367

[162] Ibid. p. 368

[163] Organization of American States. 2013. *Inter-American Convention Against All Forms of Discrimination and Intolerance*. Organization of American States, available at: https://www.oas.org/en/sla/dil/inter_american_treaties_a-69_discrimination_intolerance.asp

freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.".[164] However, in paragraph 2, the article specifies that:

> The exercise of these freedoms, since it carries with it duties and responsibilities may be subject to such formalities, conditions, restrictions, or penalties as are prescribed by law and are necessary in a democratic society […] for the protection of health or morals, for the protection of the reputation or rights of others […].[165]

Thus, the convention provides for limitation to the right to freedom of expression in some instances, allowing for hate speech cases to be brought to the court. Additionally, article 14 of the convention, the anti-discrimination article, states that the provisions contained in the convention are to be applied without discrimination on any ground and article 17, on abuse of rights, prohibits the abuse and states that nothing in the convention may be interpreted as justifying or implying that any individual, group or state party can perform an act that aims at the destruction of any of the rights included in the convention, nor can they limit rights to a larger extent than what the convention provides.[166] It is worth noticing that, although the convention contains the guarantee of freedom of expression, the prohibition of discrimination and the abuse of rights clause, it does not, contrary to some of the previously analyzed international instruments, provide a specific provision that explicitly prohibits certain types of expressions that can amount to hate speech (propaganda for war, advocacy for hatred, incitement to violence).[167]

In the factsheet the court published in June 2020 regarding hate speech, it is explained how the court generally takes uses two approaches when dealing with such cases, provided for by the convention.[168] The First approach, a broader kind of approach, analyzes the case based on article 17 on the abuse of right, and entails the loss of the right to rely on article 10. The second approach, the narrow one, relies on article 10 paragraph 2 and generally evaluates the kind of restrictions that have been posed to the right to free speech, balancing the right and the legitimacy of restrictions as provided in the article.[169] In general, the Court confirmed the particular status of freedom of expression in the context of the European Convention, as it is one of the roots of democracy itself: "*Freedom of expression constitutes*

---

[164] Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14*, 4 November 1950, ETS 5, available at: https://www.echr.coe.int/documents/convention_eng.pdf
[165] Ibidem
[166] Ibidem
[167] See ICCPR art. 20
[168] European Court of Human Rights, *Factsheet - Hate speech*, Press Unit, June 2020
[169] Casarosa, F. *Handbook on Techniques of Judicial Interaction in the Application of the EU Charter – Freedom of Expression and Countering Hate Speech,* European University Institute, Robert Shuman Centre, p. 21
European Court of Human Rights, *Factsheet - Hate speech*, Press Unit, June 2020, p. 1

one of the essential foundations of such a society, one of the basic conditions for its progress and for the development of every man".[170] Moreover, the application of article 10 of the Convention is very broad, as it applies to everyone, includes freedom of speech and freedom to receive and give information, and is applicable to all ideas and not only those that are favorably received in the society.[171]

The ECtHR has established sets of different evaluation criteria based on the approach. In the broad approach, used in some of the more recent rulings, the Court has affirmed the inadmissibility of claims in which the individual was not entitled to the protection of article 10, based on the evaluation of the expressions made and the conflict with the underlying values of the charter. In other words, the article covers all the rights that, when invoked in particular situations, would allow individuals to derive the right to engage in behaviors or actions that would endanger or destroy the rights and freedom of the convention.[172] For example, in the case M'Bala M'Bala vs France (2015), the applicant was convicted for public insults directed, in this case, to the Jewish community. The French comedian invited to one of his shows a negationist academic, Robert Faurisson, who had been previously convicted for his denial of the holocaust and of gas chambers. When joined on stage, Mr. M'Bala M'Bala awarded the guest of a "prize" delivered to him by an actor wearing "*a pair of striped pyjamas with a stitched-on yellow star bearing the word 'Jew' – who thus played the part of a Jewish deportee in a concentration camp*.".[173] The Court affirmed that the circumstances did not qualify as satirical nor provocative: "*[…] during the offending scene the performance could no longer be seen as entertainment but had taken on the appearance of a political meeting.*" And "*through the key position given to Robert Faurisson's appearance and the degrading portrayal of Jewish deportation victims faced with a man who denied their extermination, the Court saw a demonstration of hatred and anti-Semitism and support for Holocaust denial*.".[174] The court concluded that the facts in question did not fall under the protection of article 10 of the convention and were instead an example of abuse of rights that made the claim inadmissible according to article 17 of the ECHR.[175]

Another example of the broad approach is Belkacem v. Belgium (2017), in which the court addressed the conviction of Fouad Belkacem, the leader of the organization "Sharia4Belgium", for having posted public videos on YouTube in which he called on viewers to overpower non-Muslim people and to fight them, amounting to incitement to discrimination, hatred and violence.[176] The applicant

---

[170] Handyside V. The United Kingdom, App. no. 5493/72 (ECtHR, 7 December 1976), para. 49
[171] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 26
[172] Ibid. p. 28
[173] M'Bala M'Bala v. France, App. No. 25239/13 (ECtHR, 10 November 2015), para. 5
[174] Ibid. para. 39
[175] Casarosa, F. *Handbook on Techniques of Judicial Interaction in the Application of the EU Charter – Freedom of Expression and Countering Hate Speech,* European University Institute, Robert Shuman Centre, p. 22
[176] Belkacem v. Belgium, App. No. 34367/14 (ECtHR, 27 June 2017), para. 1

argued that he did not intend to incite others to violence or discrimination and was only trying to express his ideas and opinions, therefore protected by article 10 of the ECHR.[177] Nonetheless, the court declared the application manifestly ill-founded under article 17 of the convention, considering the words of the applicant as intended to incite hatred and violence and therefore, his application was considered a case of abuse of rights that implied the impossibility for the statement to be protected under article 10. It is important to note that this kind of approach to hate speech cases has been preferably used by the court for revisionist and negationist speech, as its application seems to provide a stronger protection for victims in comparison to the application of article 10 (2).[178] Moreover, the Court used article 17 especially when confronted with hate speech forms that are not explicitly cited in article 10 of the Convention.[179]

The narrow approach, based on the restrictions included in article 10 paragraph 2 deemed necessary in the interest of national security, public safety, protection of health and morals and of the rights of others, has been used by the court for the cases in which the speech act was not considered as a threat to the fundamental values of the Convention, but in which the applicant convicted for speech acts allege a violation of article 10. The court has stressed that freedom of expression is not absolute, and it does imply exceptions. However, the restrictions must be prescribed by law, pursue a legitimate aim and be necessary in a democratic society.[180] When faced with this type of cases, the court considers the particular circumstances of the case, focusing its analysis on the central question of the purpose the applicant pursued, the aim the speech act had. If answered, this question would allow to distinguish forms of expression that are protected by the convention, even if they are shocking or offensive, and hate speech.[181] Moreover, the court focuses on the context of the expression by analyzing the presence of political discourse on the matter, which is considered necessary and generally means the Court will be less prone to justify the restriction on speech, the applicant's role in the society, the status of those targeted by the speech act, the potential impact of the expression, and the seriousness and necessity of the sanctions.[182]

In the case of Jersild v. Denmark (1994), a danish journalist was convicted for having made a documentary in which he included some extracts from an interview he conducted with a group called "the Greenjackets" which contained derogatory remarks about immigrants in the country.[183] The court addressed the issue of the balancing of freedom of expression and hate speech laws. In particular,

---

[177] Ibid. para 26

[178] Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, p. 224

[179] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 29

[180] Ibid. p. 36 – 37

[181] Ibid. p. 39.

[182] Ibid. p. 39 – 45

[183] Jersild V. Denmark, App. No. 15890/89 (ECtHR, 23 September 1994), para. 10 – 11

seen as the applicant was a journalist and was convicted for "assisting in the dissemination" of objectionable statements, the court underlined the importance of the press in providing information and its duty to do it, as well as the right of the public to receive it.[184] Even after noting that audiovisual media tend to have a more immediate and powerful effect, the court made a clear distinction between the members of the group and the applicant, who exposed the remarks they made to analyze and explain a phenomenon, and whose analysis was inserted in a news program which had referred to the public debate on the matter prior to the broadcasting.[185] Therefore, the court considered the conviction of the journalist in violation of article 10.[186]

Another example is Vejdeland and Others v. Sweden (2012), in which the court was faced with homophobic hate speech disseminated in form of leaflets in high schools by the "National Youth". The leaflets contained phrases such as "the homosexual lobby organizations are trying to play down pedophilia", homosexuality was defined as a "*a deviant sexual proclivity with a morally destructive effect on the substance of society*".[187] This was the first case involving homophobic hate speech at the court and it seems that the court used the opportunity to close some of the gaps in European legislation by underlying that this type of discrimination is as serious as racial and religious discrimination are. Moreover, the court underlined that:

> […] inciting to hatred does not necessarily entail a call for an act of violence, or other criminal acts. Attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favor combating racist speech in the face of freedom of expression exercised in an irresponsible manner.[188]

For these reasons the court found no violation of article 10 by Sweden. By analyzing this decision it is possible to affirm that in the ECtHR case law, the threshold for hate speech is lower than in the EU framework decision on racism seen as the speech act is not necessarily a clear call for violence to be considered punishable. The threshold in this case shows the recognition of the harm that speech itself can create, and is closer to article 4 of the ICERD, in that dissemination of racist or discriminatory ideas in itself can constitute a punishable act.[189]

---

[184] Jersild V. Denmark, App. No. 15890/89 (ECtHR, 23 September 1994), para. 35

[185] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 47

[186] Casarosa, F. *Handbook on Techniques of Judicial Interaction in the Application of the EU Charter – Freedom of Expression and Countering Hate Speech,* European University Institute, Robert Shuman Centre, p. 22

[187] Vedjdeland and Others v. Sweden, App. No. 1813/07 (ECtHR, 9 February 2012), para. 8

[188] Ibid. para. 55

[189] Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, pp. 223 - 224

However, a more detailed explanation of the balancing between freedom of expression and the limits it can be submitted to can be found in the case of Pernicek v. Switzerland (2015). The case concerned the criminal conviction of a Turkish politician who, in May 2005, affirmed in a press conference that the deportation and the violence that the Armenian population suffered at the beginning of the 20th century did not amount to genocide. In this case, the Court analyzed article 10 and article 8 of the convention, which concerned the right to respect of private life of the Armenians, to carry out a balancing of the two rights, and affirmed, based on previous case-law, international law and relevant national law:

> that the applicant's statements bore on a matter of public interest and did not amount to a call for hatred or intolerance, that the context in which they were made was not marked by heightened tensions or special historical overtones in Switzerland, that the statements cannot be regarded as affecting the dignity of the members of the Armenian community to the point of requiring a criminal-law response in Switzerland, that there is no international-law obligation for Switzerland to criminalize such statements, that the Swiss courts appear to have censured the applicant for voicing an opinion that diverged from the established ones in Switzerland, and that the interference took the serious form of a criminal conviction.[190]

Therefore, the conviction of the applicant was considered a violation of article 10 of the Convention as it was not necessary, in a democratic society, to limit this type of expression in favor of the rights of the Armenian community that were relevant in this case.

1.4.1 Inconsistencies in the jurisprudence

As stated in the previous paragraph, the Court generally applies article 17 in the most severe cases, namely cases that deal with negationist or revisionist speech. However, there have been some interesting exceptions in this regard. For example, in the case of Norwood v. The United Kingdom (2004), the Court considered the application inadmissible under article 17 of the Charter. The applicant was a regional organizer of the British National Party, a far-right party in the country, and in the last months of 2001 he displayed in the window of his flat a poster supplied by the party that portrayed the Twin Towers burning after the terrorist attack of 9/11 with the words "Islam out of Britain – Protect the British People".[191] The applicant was charged and convicted of aggravated

[190] Perinçek V. Switzerland, App. No. 27510/08 (ECtHR, 15 October 2015), para. 280
[191] Norwood v. The United Kingdom, App. No. 23131/03 (ECtHR, 16 November 2004), p. 2

offence for the displaying of hostility towards racial or religious groups in 2002, he then appealed to the court in 2003 for the alleged breach of article 10.[192]

In this case the Court referred to article 17 of the convention, stating that the poster and the words written in it amounted to a public attack to the Muslims in the country and that "*Such a general, vehement attack against a religious group, linking the group as a whole with a grave act of terrorism, is incompatible with the values proclaimed and guaranteed by the Convention.*". [193] However, contrary to previous case-law, the case did not fall into the severity threshold that usually accompanied the use of article 17 and it is unclear why the Court decided that this reasoning was to be applied in this case, while later similar cases were judged through article 10 of the convention. In Vejdeland and Others v. Sweden for example, the distribution of leaflets with harsh homophobic remarks was considered under article 10, and not article 17.[194] Similarly, in Balsyte-Lideikiene v. Lithuania (2009), concerning the publication of a calendar that contained statements that incited to hatred towards Jews and Polish people, was approached referring to article 10 and the measures adopted by the government were considered necessary in a democratic society.[195]

A partial explanation of the threshold used in the application of article 17 can be found in the case of Soulas and Others v. France (2008). The case concerned a proceeding against the applicants for the publication of a book titled "The colonization of Europe" that contained statements inciting hatred and violence against the Muslim communities.[196] In this instance, the court considered there had been no violation of article 10 by the French courts and gave an explanation on why article 17 was not applied. France invoked article 17 of the Convention and affirmed that the application should have been deemed inadmissible seen as the depiction of Muslims in the publication was founded on racist and discriminatory consideration that, in their view, made the applicant's appeal to freedom of expression as an abuse of rights, going against the principles of the Convention itself. However, the Court observed that the passages contained in the publication, although sanctionable under article 10 paragraph 2 of the convention, were not severe enough to justify the application of article 17.[197] By this reasoning, it is possible to state that the court "*has embraced a relatively low threshold in finding hate speech*",[198] in the sense that the expression of racial hatred is, for what concerns the case-law of the ECtHR, outside the protection of article 10 paragraph 1.

---

[192] Ibidem.
[193] Ibid. p. 4
[194] Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, p. 225
[195] Ibid. p. 224
[196] Soulas v. France, App. No. 15948/03 (ECtHR, 10 October 2008), para. 6 – 7
[197] Ibid. para 20
[198] Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, p. 225

1.5 Interamerican Court of Human Rights Jurisprudence on Hate Speech

The American Convention on Human Rights was elaborated almost twenty years after the European counterpart, in 1969, and established the limits and scope of the regulation of freedom of expression in article 13. The InterAmerican convention explicitly recognizes the right to seek information in the first paragraph of article 13 stating:

> Everyone has the right to freedom of thought and expression. This right includes freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing, in print, in the form of art, or through any other medium of one's choice.[199]

The Convention then states in paragraph 2 that the exercise of freedom of expression can be limited only to ensure the respect of the rights and reputation of others, other than to protect national security, public health, and morals and only if the restrictions are expressly established by law and necessary. Moreover, the document expressly prohibits prior censorship, the banning of any type of expression before it is produced, with the only exception being regulating access to contents for the moral protection of children or adolescents. Lastly, similarly to other international conventions, article 13 paragraph 5 provides that:

> […] any propaganda for war and any call to hatred of national, racial or religious hatred that constitutes incitement to lawless violence or to any other similar action against any person or group of people for any reason, including because of race, colour, religion, language or national or social origin shall be considered as offenses punishable by law.[200]

The courts approach to freedom of expression is, as for the ECtHR, to privilege the extension of the right rather than its limitation, therefore limitations must be restrictively interpreted.[201] Contrary to the European counterpart, the American Court of Human Rights has a limited case law in this regard. However, it did provide explanatory interpretations of the articles of the convention in some advisory opinions.

In "Compulsory Membership in an Association Prescribed by Law for the practice of Journalism" (1985) the court addressed the absence of the expression "just demands of a

---

[199] Organization of American States (OAS), *American Convention on Human Rights, "Pact of San Jose", Costa Rica*, 22 November 1969, available at: https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm

[200] Ibidem

[201] Úbeda de Torres, Amaya. "Freedom of Expression under the European Convention on Human Rights: A Comparison with the Inter-American System of Protection of Human Rights." Human Rights Brief 10, no. 2 (2003). p. 7

democratic society" as a clause for the limitation of freedom of expression.[202] The case concerned the request for an advisory opinion brought by the Costa Rican Government to the Court for the interpretation of article 13 in relation to the compulsory membership to the journalism association created by the government in 1969. In the opinion, the Court referred to article 32 paragraph 2, which allows limiting rights to protect the general welfare of a democratic society. However, the court affirms that article 32 applies only when the Convention has not established legitimate restrictions on a right, which makes it inapplicable for the right to freedom of expression as its limits are already explicitly stated in article 13.[203] In a different advisory opinion, the court provided an interpretation of article 30:

> The restrictions that, pursuant to this Convention, may be placed on the enjoyment or exercise of the rights or freedoms recognized herein may not be applied except in accordance with laws enacted for reasons of general interest and in accordance with the purpose for which such restrictions have been established.[204]

In particular, the court provided a definition of the word "laws" contained in the article. The court concluded that the word means "*general legal norm tied to the general welfare, passed by democratically elected legislative bodies established by the Constitution […]*".[205]By this reasoning, seen as the article refers to the entirety of the rights and freedoms contained in the convention and to their limitation, it is possible to affirm that the rights included in article 13 must conform to the needs of the democratic society, even if it is not expressly included as a limitation.[206] In general, the Court considers that restriction to freedom of expression must meet several requirements: must be prescribed by law in a clear and unequivocal manner, must be absolutely necessary in that they must be justified by collective needs that overtake the right freedom of expression while at the same time interfering as little as possible with the enjoyment of said right, and, finally, the State must prove that the restriction was absolutely necessary to protect the rights of a third party.[207] Finally, for what concerns hate speech specifically, the expression "incitement to lawless violence or any other similar

---

[202] Global Freedom of Expression website: https://globalfreedomofexpression.columbia.edu/cases/la-colegiacion-obligatoria-de-periodistas-oc-0585/ [Accessed 8 December 2021]
[203] Ibidem
[204] Organization of American States (OAS), *American Convention on Human Rights, "Pact of San Jose", Costa Rica*, 22 November 1969, available at: https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm
[205] *Advisory Opinion OC-25/18 OF 30 May 2018 Requested By The Republic Of Ecuador*, Inter-American Court of Human Rights (IACrtHR), 30 May 2018, p. 9
[206] Úbeda de Torres, Amaya. 2003,"Freedom of Expression under the European Convention on Human Rights: A Comparison With the Inter-American System of Protection of Human Rights." *Human Rights Brief* 10, no. 2, p. 8
[207] Posenato, Naiara. "The Protection Of The Right To Freedom Of Expression: A Panorama Of The Inter-American Court Of Human Rights Case Law", Espaço Juridico Jurnal of Law (EJJL), v. 16 (2015), Edição Especial - Graves Violações DDHH, p. 63

action” contained in article 13 (5) suggest that in the American system a necessary requirement for speech to be considered hate speech is violence, which makes the article substantially different from other similar documents in that it does not include discrimination or hostility as possible characteristics of hate speech.

Chapter 2
Controversies in Hate Speech regulation

2.1 Freedom of Speech and Hate Speech: Balancing approach problems

In the previous chapter the focus was on the norms and standards for hate speech in international law; however, through the analysis of the existing legislation, some issues and controversies become evident. It is possible to notice how the concept alone arises some concerns, as the lack of a universally accepted definition of the term results in a lack of clarity, and scholars and law makers do not agree on what hate speech is, on whether it should be banned, on how the bans should be enacted.[208]

The term has been included in many state's legislations, but the definitions provided are slightly different, which creates a difficulty in identifying which speech acts can be considered as hate speech.[209] For example, the Dutch criminal code calls for the punishment of anyone who incites hatred or discrimination against people or groups of people because of their race, religion, sex, sexual orientation, or disability.[210] However, the Dutch definition does not include, for example, the public condoning, denial or trivialization of international crimes and the Holocaust that are required under the Council Framework Decision of the EU.

Moreover, the concept of hate speech and its criminalization is not universally accepted as legitimate, and scholars are still divided on the matter. According to some scholars, "*[h]ate speech undermines this public good, or it makes the task of sustaining it much more difficult than it would otherwise be*"[211], compromising the dignity of the targets and their reputation by using personal characteristics as disqualifying attributes.[212] For others the nature of the bans on speech amount to censorship and are to be avoided at any cost because "*[w]ithout freedom of speech and the right to dissent, the Civil Rights movement would have been a bird without wings*".[213]

2.1.1 The absence of a broadly accepted definition and clear standards for Hate Speech

The proliferation of hate speech legislation has been occurring since World War II. Nevertheless, as mentioned in the previous chapter, the term hate speech does not have a clear definition in

[208] Carlson, C. R. 2021, *Hate Speech*, Cambridge, The MIT Press, p.12
[209] Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing, p. 3
[210] Dutch Government website: https://wetten.overheid.nl/BWBR0001854/2020-07-25/#BoekTweede_TiteldeelV_Artikel137d [Accessed 8 January 2022]
[211] Waldron, J. 2012, *The Harm In Hate Speech*, Cambridge, London, Harvard University Press, p. 4
[212] Ibid. p. 5
[213] Congressman John Lewis in Strossen, N. 2018, *HATE - Why We Should Resist It with Free Speech, Not Censorship,* Oxford, Oxford University Press, p. 110

international law, to the point where the term alone arises some criticism. Scholars have argued that the use of the word "hate" to define the type of speech prohibited under international human rights law is improper. Hate is a strong emotion, and the connotation of this type of speech as connected to a subjective mood such as hate can limit the understanding and the application of the concept, as some examples of hate speech do not come from an emotional and unreasoned state, but take the form of a calculated and pseudo-scientific argument.[214] Another argument that arises from the presence of many different definitions of hate speech in the international conventions and regional laws is that this situation produces conflicting standards and norms by contributing to a general confusion about the possible limits of speech.[215] Although the definitions we find in the law overlap significantly, they do not create a consistent framework. Even in the same regional context, definitions and standards differ. In the European context, for example, the ECHR views as illegal types of expression those that may be insulting to individuals or groups and the expressions that spread, promote, or justify hatred.[216] However, the Additional Protocol to the Convention on Cybercrime calls for an even broader prohibition of threats, public insults and dissemination of ideas based on racial superiority.[217] Instead, Council framework decision 2008/913/JHA requires a higher threshold for the criminalization of forms of expression: intent and incitement to hatred.[218] By comparing the provisions to the other international law sources, other differences can be highlighted: the standards of the Framework Decision are almost entirely aligned with the Article 20(2) of the ICCPR, but intent is not mentioned in the international convention.[219] In the CERD, states are obliged to "*declare an offence punishable by law all dissemination of ideas based on racial superiority*"[220], which is once again different from the prohibition of "*any advocacy of national, racial or religious hatred*"[221] contained in the ICCPR. According to Jacob Mchangama, a Danish human rights advocate, this lack of a coherent and unified approach to the legislation and the consequent cases of hate speech can

---

[214] Heinze, E. 2016, *Hate Speech and Democratic Citizenship*, Oxford, Oxford University Press, p. 22
[215] Mchangama, Jacob. 2015, "THE PROBLEM WITH HATE SPEECH LAWS", *The Review of Faith & International Affairs*, vol. 13 , pp. 77
[216] Gündüz v. Turkey, App. No. 35071/97 (ECtHR, 4 December 2003), para 40
[217] Council of Europe, Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, European Treaty Series, No. 189, Strasbourg, 28.I.2003
[218] Council of the European Union, FRAMEWORK DECISION 2008/913/JHA, on combating certain forms and expressions of racism and xenophobia by means of criminal law, L 328/55
[219] Mchangama, Jacob. 2015, "THE PROBLEM WITH HATE SPEECH LAWS", p. 78
[220] UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, p. 195, available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx
[221] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

result, and has resulted, in problematic convictions and abuses, with hate speech laws used as justifying tools to delegitimize and prosecute legitimate political speech.[222]

As a result of these difficulties in identifying the limits and the scope of the provisions on hates speech, the approach that states, especially liberal democracies, have to the issue is still very heterogenous. If European democracies have been enacting laws on the matter, trying to promote equality while limiting the right to freedom of expression, the United States have taken a completely opposing route, affording protection to hate speech under the First Amendment.[223] These two opposing approaches stem from a disagreement over the rights considered as most important by the different traditions. As a result, even within the western democracies there is a lack of homogeneity in the definition and scope of hate speech legislation. The United States Supreme Court has repeatedly stated that the offensiveness of speech does not justify a suppression of the speech act, as there is no constitutional basis for the prohibition and the suppression creates a danger to democracy.[224] Moreover, even within European countries the historical experiences and legal tradition plays a central role in defining the scope of the provisions on hate speech, making the laws on the matter substantially differ from one another. In Germany, the constitution states that human dignity is inviolable and therefore is preeminent in the constitutional order.[225] For this reason, Germany's high court recognized limitations to freedom of speech as constitutional and held that the state has a positive obligation to protect equality and dignity. This sensitivity towards dignity and equality is a result of the historical context of Germany and the Nazi experience, which makes German hate speech laws some of the most extensive, as the criminal code prohibits "*incit[ing] hatred against segments of the population or call[ing] for violent or arbitrary measures against them; or* **...** *assault[ing] the human dignity of others* ***by*** *insulting, maliciously maligning or defaming segments of the population*".[226]

In the United Kingdom, the Race Relations Act of 1965 criminalized threatening, abusive and insulting speech as well as speech that "*intended to incite hatred on the basis of race, color or national origin*".[227] In 1986, the parliament added a provision in the Public Order Act that prohibits any action intended to incite racial hatred, and in 2006 the Racial and Religious Hatred Act was adopted. With

---

[222] Mchangama, J. 2015, "THE PROBLEM WITH HATE SPEECH LAWS", *The Review of Faith & International Affairs*, vol. 13, p. 78

[223] Cohen, Roni. (2014) "Regulating Hate Speech: Nothing Customary about It" *Chicago Journal of International Law*: Vol. 15: No. 1, Article 11, p. 231

[224] Ibid. p. 245

[225] German Federal Ministry of the Justice website: https://www.gesetze-im-internet.de/englisch_gg/ [Accessed 9 January 2022]

[226] Cohen, Roni. (2014) "Regulating Hate Speech: Nothing Customary about It" *Chicago Journal of International Law*: Vol. 15: No. 1, Article 11, p. 240

[227] UK Parliament website: https://www.parliament.uk/about/living-heritage/transformingsociety/private-lives/relationships/collections1/race-relations-act-1965/race-relations-act-1965/ [Accessed 9 January]

these three provisions, the UK criminalizes the "*promotion of hatred through persuasion of non-target audiences [...] if it amounted to harassment of a target group or individual*".[228]
The identification of what constitutes hate speech is usually resolved in law by referring to the category of incitement, which effectively identifies the more evident forms of hate speech, what Yong identifies as "targeted vilification" and "diffuse vilification", while more covert forms of hate speech are ignored, as there has been little legal or philosophical discussion on what Yong defines as "*organized political advocacy for exclusionary policies*" and "*other assertions of fact or value which constitute an adverse judgment on an identifiable racial or religious group*".[229] In fact, the term "incitement" is used to identify the more evident forms of discrimination, which are more likely to provoke violence, while ensuring a certain protection to the right to free speech, as the emphasis does not fall on the speech but on the possible consequences.[230] However, this approach to the legal categorization of hate speech can lead to an incorrect categorization of certain forms of speech. Speech acts that are "*expressed in a civil or reasonable way or hate speech which has some bearing on a political issue or issue of public concern*"[231] is often overlooked and seen as mere academic debate or political debate, thus justified with the "search for truth" argument or the democratic argument. The confusion between the two concepts, incitement and hate speech, creates a dangerous situation in which a speech act could be considered either a terrible crime (incitement to genocide, which is criminalized in international treaties and most national laws) or as the legitimate exercise of a right.[232]
In her article "Hate Speech and Distorted Communication" Sarah Sorial provides two concrete examples of this issue, one from the Canadian Human Rights Tribunal (R v Zundel) and one from the Australian Federal Court (Jones v Toben). The two cases related to Holocaust denial on Internet websites resulted in the prosecution of the two authors. However, according to the author of the paper, although the two were prosecuted, the decision of the two courts did not stem from the contents of the websites; the courts decided as they did because of the inflammatory language used by the authors. "*Had the views been expressed in more measured and reasonable language, the speech would have constituted legitimate academic debate, and Zundel and Toben may have escaped prosecution*".[233] In fact, in the case of R v Zundel, faced with evidence stating that the materials on the website were a

---

[228] Rosenfeld, Michel. 2003, "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis ", *Cardozo Law Review*, p. 1547
[229] Yong, Caleb. 2011, "Does freedom of speech include hate speech?", *Res Publica*, vol. 17, no. 4, p. 385
[230] Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, p. 300
[231] Ibid. p. 303
[232] Benesch, Susan, 2008, "Vile Crime or Inalienable Right: Defining Incitement to Genocide", *Virginia Jurnal of International Law*, Vol 48:3, p. 493
[233] Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, p. 304

legitimate expression of historical debate, the court seemed to accept this reasoning, stating "*Our conclusion is based on the way in which these doubts are expressed, and not on the fact that challenges are raised regarding the historical accuracy of these events*".[234] Similarly, in the Australian case, the court argued that the issue in the expression arose from the lack of effort by the author in expressing his views with constraint and restraint. The result is that the speech act was considered as hate speech not for its contents but because of how the views were expressed, since the delivery of the speech could, according to the court, be considered incitement. This understanding of hate speech is concerning for two main reasons: first, the speech (be it academic or not) is directed at groups that have already been oppressed by the society. The speech might enhance the perception of those groups as less than equals in the society, and therefore affecting the dignity of the group. Second, the justification of hate speech disguised as academic debate might aid in making the views expressed appear as acceptable, realistic, and appealing.[235]

2.1.2 The ideological debate: is the limitation to free speech acceptable?

One of the main issues with the affirmation of the concept of hate speech and its criminalization is the debate on the legitimacy of speech limitations. In general, free speech is considered vital in every aspect of the public and private life. Free speech is the central pillar of thought and critical consciousness, "*When it is denied or severely curtailed, the human capacity to think, and all that is distinctive to human beings, is undermined*".[236] Moreover, because speech is necessary to live a meaningful human life, if speech is significantly limited and constrained, the relations and bonds between humans become fragile and lack honesty and permanence. In political life, free speech is essential to ensure that citizens can and do provide opinions and belief, that decisions undergo a real critical scrutiny, and that the government is effectively checked.[237] These are some of the reasons why the right to free speech is considered essential and is universally protected in the international community. However, the fact that speech is such an important aspect of the lives of people and of the democratic functioning of states does not cancel out the presence of other essential human rights, with the same amount of importance. Dignity, freedom from harassment and equality are also central to a meaningful life. Since these values are considered equally essential but can, in some situations, be in conflict with each other, there must be a balancing: "*Every value makes claims that limit those of others, and every right is limited in its content and scope by other rights*".[238]

---

[234] Ibidem
[235] Ibid. p. 307
[236] Herz, M. Molnar, P. 2012, *The Content and Context of Hate Speech – Rethinking Regulation and Responses*, Cambridge, Cambridge University Press, p. 42
[237] Ibid. p. 43
[238] Ibidem

To understand the characteristics of the debate the categorization applied by Erich Heinze in the book "Hate speech and democratic citizenship" provides a useful overview of the main positions on hate speech bans. The arguments regarding the banning of hate speech, according to Eric Heinze, usually fall under two types of reasoning: the consequentialist (outcome based) and the deontological (duty based). Each one of these categories can be sub-divided into two more types of reasoning: those that support bans (prohibitionism) and those that reject bans (oppositionism).[239] The consequentialist prohibitionist reasoning affirms that hateful expression can lead to harmful effects to specific target groups, and therefore punitive bans on discourse are necessary or at least permissible to combat the effects of hate speech. Deontological prohibitionist arguments come to the same conclusions but the reasoning stems from a different thesis: hateful expression degrades the target's intrinsic dignity, therefore there is no need for an evident harmful effect on the group or the person for the bans to be considered necessary or permissible. Conversely, consequentialist oppositionism rejects bans on speech because bans can call forth the exact situations and harmful effects that they are trying to prevent, making the ban counterproductive. Lastly, deontological oppositionism views the presence of punitive bans on viewpoint as a threat to the legitimacy of the democracy, thus making bans illegitimate.[240]

A common concern expressed by scholars to the regulation of hate speech is rooted in political philosophy, more specifically, in Stuart Mill's views on freedom of expression. In his work "On Liberty" Mill provides a standard for which restrictions are acceptable in a civilized community: a limitation to one's liberty, in his view, is only legitimate if it is used to avoid harm to someone else. However, what harm falls into this "harm principle" categorization is not as easy to identify, although it is clear offence and inconvenience are not considered harm.[241] The Harm principle can therefore be understood as a state's obligation to intervene to avoid violations of rights, provided that the violation falls in the millian threshold of harm. In his utilitarian view, some forms of harm should not be prevented by the state, not because of an evaluation in terms of the validity of the harm suffered, but because "*cost of the harm is, on balance, worth the benefit of increased freedom*".[242] Regarding censorship, he focuses on the censorship that could suppress a false or immoral opinion, hence it is possible to say that he refers to what is now considered hate speech. In discussing this type of censorship, he provides four reasons to protect and maintain free speech instead of banning certain forms of expression: the opinion might be true, might contain a part of truth, the censorship of a false opinion might turn true opinions into a dogma, and a dogma, as an unchallenged opinion, loses its

---

[239] Heinze, E. 2016, *Hate Speech and Democratic Citizenship*, Oxford, Oxford University Press, p. 33
[240] Ibid. p. 34
[241] Brink, O. David. 2001, "Millian Principles, Freedom of Expression, and Hate Speech", *Legal Theory,* Vol. 7, p. 121
[242] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 14

meaning.[243] This reasoning directly connects to the marketplace of ideas theory by the same author, according to which the most truthful and persuasive ideas are those that can withstand the pressure of the open debate. Consequently, ignorant ideas - those that do not promote truth - do not need to be prohibited or punished, as the scrutiny that these ideas will eventually undergo in the public debate will lead to their discarding.[244]

A more recent contribution to the debate comes from Ronald Dworkin, an American philosopher, who justified freedom of expression based on equality. The argument he makes is that freedom of expression is "*absolutely crucial to moral agency, and that moral agency is the cornerstone of democratic culture*".[245] For this reason, individuals should have the same opportunity and possibility to influence our environment in moral terms. Therefore, the involvement of state in one's right to freely express their ideas is, according to the scholar, a violation of the commitment of states to equality. Dworkin's view differs from Mills not only because of the justification of the right to freedom of expression (equality instead of liberty) but also because of the different approach used in identifying the value of the right. If Mill provided instrumental justifications in a utilitarian way, Dworkin tries to capture the real, constitutive value of freedom of speech. "*For a sustainable democratic culture, it is necessary both that individuals are independent moral agents (or at least have the inherent potential to develop into them), and that government treat them as such*".[246] Moral independence is a necessary requirement of a democratic culture, and moreover, Dworkin argues that freedom of expression is a constitutive element of such moral independence.[247]

This reasoning provides a strong argument for an almost absolute right to free speech, but prompts extensive criticism, especially by those who believe hate speech is problematic for the moral agency of minorities and discriminated groups. Critical race theorists and feminist theorists have argued, referring primarily to the debate regarding the US approach to the issue but more in general to Dworkin's theory, that the doctrine of free speech has developed without considering equality in a serious manner. According to Catharine MacKinnon, Rae Langton and Jennifer Hornsby, hate speech is a danger to the equality principle because the views expressed in hate speech acts aid in the rising and spreading of unequal opportunities for the minorities involved. In other words, hate speech acts enact subordination and therefore are a threat to equality.[248] As quoted by Abigail Levin in the book "The Cost of Free Speech":

---

[243] Brink, O. David. 2001, "Millian Principles, Freedom of Expression, and Hate Speech", *Legal Theory,* Vol. 7, p. 122
[244] Heinze, Eric. "Viewpoint Absolutism and Hate Speech." *The Modern Law Review*, vol. 69, no. 4, 2006, p. 554
[245] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 81
[246] Ibid. p. 82
[247] Levin, Abigail. 2009, "Pornography, Hate Speech, and Their Challenge to Dworkin's Egalitarian Liberalism", *Public Affairs Quarterly*, Vol. 23 No. 4, p. 357
[248] Ibid. p. 360

> [T]he First Amendment has grown as if a commitment to speech were no part of a commitment to equality […] Understanding that there is a relationship between these two issues – the less speech you have, the more the speech of those who have it keeps you unequal; the more the speech of the dominant is protected, the more dominant they become and the less the subordinated are heard from – is virtually nonexistent.[249]

This argument, the subordination argument, is a particularly effective countering of Dworkin's theory on freedom of expression. In fact, Dworkin asserts that absolute freedom of expression is protective of equality and a necessary condition for the existence of a democratic community based on moral membership. However, Mackinnon's counterargument entails that freedom of expression is, in some cases, not only untied from equality, but even damaging it.[250]
Another effective argument countering unregulated freedom of expression is the silencing argument. This argument is mostly applied to the issue of pornography but does provide an interesting prospective on the effects of unregulated speech. The central thesis of this argument is that hate speech and pornography systematically silence the subsequent speech of women and minorities for many different reasons: fear and cynicism might prevent them from even trying to respond, and even in the case a response is made, the context in which it is inserted will probably result in the rebuttal being ignored and misunderstood by the majority.[251]

> Thus, while, of course, women and minorities are still tech anyone else to speak, the silencing argument holds that the background conditions for their speech, having been established by the preceding racist or sexist speech, are such that any subsequent speech is discounted in advance by the privileged recipients, or not spoken at all by the oppressed speakers.[252]

To summarize, the two arguments are denying the presence of a constitutive connection between freedom of expression and human dignity and equality by asserting that: "[*I]t seems that freedom of expression cannot be constitutively tied to dignity and moral agency if in some cases the exercise of freedom of expression denies the dignity and moral agency of its targets".*[253]

---

[249] Mackinnon in Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan
[250] Levin, Abigail. 2009, "Pornography, Hate Speech, and Their Challenge to Dworkin's Egalitarian Liberalism", *Public Affairs Quarterly*, Vol. 23 No. 4, p. 361
[251] Ibid, p. 362
[252]Ibid. p. 363
[253] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 91

Another interesting contribution to de debate comes from J. L. Austin speech acts theory. In the book "How to do things with words", Austin tries to revise the different functions of speech acts and describes three kinds of speech acts and their central differences. First, locutionary speech acts, which seek to describe the state of the world, are the most common and the words are a mere instrument to describe something after it has happened, are defined as factual statements and categorized as constative speech.[254] Although these might seem like the only type of speech acts, the author argues that there are two more types of speech acts which, instead of being passive recipients of previous facts, directly influence the conditions in the world after being expressed, categorized by the author as performative speech acts.[255] The second type of speech acts, perlocutionary utterances, "*create effects in the world not simultaneously with their utterance, but after the recipient or listener performs a mental act which concludes with her taking up a different stance or position than she occupied before the utterance*".[256] In this case, the speech does not have a truth value but is dependent on the subsequent measurement and mental act by the recipient, and therefore speech is actively participating in the subsequent events.[257] The third possible speech act, the most relevant in the context of hate speech theory, is an illocutionary speech act. In this case, the speech is doing things directly, instead of reporting a fact or depending on the listener's reaction. For example, the statement "I do" during a wedding is an act that alone creates consequences, the marriage. However, for this type of speech to be effective, there must be some background conditions: there must be an accepted conventional procedure that provides meaning, the circumstances must be appropriate, and the procedure must be executed correctly. Thus, saying "I do" in the marriage ceremony would be useless if, for example, the person celebrating the wedding had no legal authority to do so.[258] Therefore, to analyze a performative illocutionary speech act the entirety of the speech situation and context needs to be taken into consideration.

This approach to speech acts which goes beyond the meaning of the spoken words and considers the entirety of the concrete context and social facts that accompany the speech act has important implications on the debate on freedom of expression, as it provides a justification for evaluating speech not only by considering the words used but also in relation to the overall circumstances in which they are inserted.[259] Starting from this theory, Rae Langton provides a further category of speech acts, authoritative illocutions, which are essentially illocutionary acts that rely not only on the

---

[254] Langton, Rae. 1993, "Speech Acts and Unspeakable Acts", *Philosophy and Public Affairs*, Vol. 22, No. 4. p. 295
[255] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 106
[256] Ibidem
[257] Langton, Rae. 1993, "Speech Acts and Unspeakable Acts", *Philosophy and Public Affairs*, Vol. 22, No. 4. p. 296
[258] Ibid. p. 496
[259] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 109

social context, but also on the position of authority that allows the act to work as previously planned. "*The ability to perform speech acts of certain kinds can be a mark of political power [...] If you are powerful, there are more things you can do with your words*".[260] According to Langton, although the possibility to subordinate through speech is more common when thinking about state authority, there are other types of authority that can subordinate. In fact, the author asserts that even the actors who possess lesser authority than states can subordinate "*when either the local perceived legitimacy or the local efficacy of the utterance is sufficient, subordination is effected*",[261] which makes the assumption that the culture has the power to subordinate minorities through hate speech and other speech acts plausible since society has been historically racist, sexist and homophobic, among other things.[262]

2.2 Online Hate Speech

Starting from the 1990's, the expansion of the internet has resulted in a drastic change in all aspects of life, but especially in information transmission and in communication. Prior to the exponential growth of the internet, the public had been mostly a passive subject of information transmission, which was dominated in large part by the mass media, lobbies, and states.[263] Today's situation is entirely different: citizens have become information transmitters as well as creators of content, acting in a cyber space in which the technological mean reduces time and space limitations and allows the public to offer opinions, open dialogue and share knowledge with virtually non-existent limits. However, the online setting does create some new challenges, in that the users can discuss and express themselves without limits, often protected by varying degrees of anonymity, "*the ability to perform any access, communication or publication in the network without third parties having the possibility to identify or locate the author of said action*".[264] According to Salvador Carrasco, anonymity creates a situation that facilitates and allows criminal activities of different types to be carried out in the cyberspace, from plagiarism, to hate speech, to terrorism.[265]

In the case of Hate speech, the question of how it should be addressed on the internet becomes even more challenging. The anonymity and immediacy of the internet have proven to be ideal for extremist groups and individuals in promoting hateful ideals and activities. Over the years in which computers

[260] Langton, Rae. 1993, "Speech Acts and Unspeakable Acts", *Philosophy and Public Affairs*, Vol. 22, No. 4. p. 298 - 299

[261] Levin, A. 2010, *The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan, p. 113

[262] Ibid. p. 113

[263] Assimakopoulos, Stavros. Baider, Fabienne. Millar, Sharon. 2017, *Online Hate Speech in the European Union: A Discourse-Analytic Perspective,* Cham: Springer Open, p. 11

[264] Carrasco, Luis. 2012, "Redes De Anonimización En Internet: Cómo Funcionan Y Cuáles Son Sus Límites", *Istituto espanol de estudios estrategicos*, p. 2

[265] Ibiedm

and later cellphones have begun to rise in numbers, hate groups and websites have also grown exponentially.[266] Hate speech, harassment and discrimination have risen as fast as hate groups in the cyberspace and policing it has proven to be a challenge. To tackle the issue, many states have enacted unilateral strategies to try and impose some virtual borders onto the internet and to prosecute the dissemination of hate speech online. However, unilateral efforts are largely limited by the reach that national legislation has, far more limited than that of the material it attempts to regulate. "*A single national or institutional entity does not control the Internet, and there is a multiplicity of jurisdictions that affect the operations of global Internet companies carrying a multiverse of content*".[267] Moreover, the divergence between different national legislations and approaches results in an even more difficult application of the laws.[268]

In a landmark case, two French student organization tried to have the Internet provider Yahoo prosecuted for violating a French law that prohibits the offering for sale of Nazi memorabilia. The internet provider argued that, since the content had been uploaded in the US where the conduct is not illegal, the French court did not have jurisdiction. The court ruled against this argument and requested that the company used a mechanism to reduce access to such merchandize by French citizens within three months to avoid a fine.[269] However, the company took the case to the United States District Court, to obtain a court ruling stating that the French ruling was a breach of the First Amendment and obtained such ruling.

> Although France has the sovereign right to regulate what speech is permissible in France, this Court may not enforce a foreign order that violates the protections of the United States Constitution by chilling protected speech that occurs simultaneously within our borders.[270]

This impasse between two national courts is the exemplification of the hardship of regulating online hate speech with national means that do not have a way to extend their reach out of the borders of the nation state.

The problem of the definition of hate speech is also enhanced by the fact that most social media platform provide their own definition of the concept and their own standards for what content that

---

[266] Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 234

[267] Ülgen, Sinan. 2016, "FREEDOM OF EXPRESSION ONLINE. GOVERNING CYBERSPACE: A Road Map for Transatlantic Leadership", *Carnegie Endowment for International Peace*, p. 17

[268] Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 234

[269] Ibid. p. 235

[270] Yahoo! Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, 169 F. Supp. 2d at

amounts to hate speech, which creates a lack of uniformity in identifying hate speech online.[271] However, even after taking steps to ensure that hateful content is regulated and removed, internet companies still seem to be inconsistent and not entirely transparent on the matter. Moreover, the amount of content that is uploaded to the platforms makes for an added level of difficulty in regulating it, as it is nearly impossible to review the thousands of videos, posts and comments that are uploaded every day. [272]

2.1.1 Characteristics of online hate speech

As previously mentioned, the conveyance of hate speech through the internet creates some unique characteristics that differentiate this branch of hateful expression from the traditional, offline hate speech. One of the characteristics that allows for the distinction between offline and online hate speech is the already mentioned anonymity. This feature of the Internet is a double-edged sword: on one side, it provides an opportunity for freer speech because people can express their opinions without being judged for their skin color, gender, or sexual orientation, as they do not have to disclose this information unless they wish to do so.[273] The presence of online forums where users have the possibility to express themselves anonymously allows for an opportunity to exchange cultural and political views and ideas, especially for citizens of repressive regimes.[274] Furthermore, the condition of anonymity can allow victims of hate speech to engage in counter speech free from the fear of being identified by hate speakers and persecuted in real life.[275]

On the other hand, the anonymity facilitates the carrying out of illegal activities, from hate speech to terrorism.[276] In a classic study by Zimbardo, students were asked to deliver electric shock to two women after having been divided into two groups. One group was given name tags and was easily identifiable, while for the other the name was replaced with a number and students were given clothing that hid their face, providing them with a sense of anonymity. The study found that the students that were "anonymous" delivered twice the amount of electric shock compared to the other group.[277] From a psychological point of view, the condition of perceived anonymity acts as a

[271] Ülgen, Sinan. 2016, "FREEDOM OF EXPRESSION ONLINE. GOVERNING CYBERSPACE: A Road Map for Transatlantic Leadership", *Carnegie Endowment for International Peace*, p. 18
[272] Ibidem
[273] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 298
[274] Carrasco, Luis. 2012, "Redes De Anonimización En Internet: Cómo Funcionan Y Cuáles Son Sus Límites", *Istituto espanol de estudios estrategicos*, p. 2
[275] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 299
[276] Carrasco, Luis. 2012, "Redes De Anonimización En Internet: Cómo Funcionan Y Cuáles Son Sus Límites", *Istituto espanol de estudios estrategicos*, p. 3
[277] Keats Citron, D. 2014, *Hate Crimes in Cyberspace*, Cambridge, London, Harvard University Press, p. 58

disinhibiting factor, encouraging users to say things that they would not dare to say in a normal face to face conversation or to act more destructively then they would if they were easily traceable.[278] Even if the condition of anonymity is not actually real (it is possible to trace the origin of online content), the perception of being untraceable and the physical separation from the target of the hateful speech makes it easy to self- justify hateful and abusive speech.[279] The lack of proximity creates a situation in which the effects of one's speech are invisible to the speaker, who is therefore operating without the "*normal social-psychological cues of empathy and censure that tend to keep harmful or antisocial behavior in check*",[280] while the sensation of anonymity is:

> regarded as a sort of "get out of jail free" card to express opinions or engage in activities the harassers know to be wrong in a social or legal sense, without dealing with any of the repercussions of having those opinions connected to their offline identity.[281]

Another characteristic of hate speech delivered through the internet is related to innate desire of people to engage with people who share the same beliefs and ideas. The internet, thanks to its accessibility and reach, is an especially effective tool in ensuring that like-minded individuals find each other on the web, and it provides the means people need to create a community.[282] Although this can be seen as a strength of the digital space, as it allows people to connect with individuals from different parts of the world, for instance members of diasporas that are able maintain contacts with their ethnic identities; when combined with the hate speech phenomenon it produces an exacerbation of the gravity of the discourse and to a radicalization of ideals. In fact, it is a common group dynamic that when people with similar ideas are united in a group, the individual will tend to embrace a more radical view on the subject as they feel more confident and want to be accepted.[283] An example of this dynamic can be found in a study conducted by Magdalena Wojcieszak based on data obtained from neo-Nazi online discussions. In this case, the results of the study suggested that the participation to this kind of homogeneous group increased members' extremism, with members in the groups encouraged each other to "*reject dissenting views encountered during daily interactions*"[284] and

---

[278] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 298
[279] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 11
[280] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 300
[281] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 23
[282] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 301
[283] Keats Citron, D. 2014, *Hate Crimes in Cyberspace*, Cambridge, London, Harvard University Press, p. 63
[284] Wojciezak, Magdalena. 2010, "'Don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism", New media and society Sage, Vol 12 No. 4, pp. 648

thorough argumentation in favor of Whit Nationalism. This polarization was observed in both long threads and short exchanges, for example: "*The White Power online forum asked participants whether they preferred extermination, segregation, or slavery as punishment for the hated group. Posters who initially opted for segregation changed their vote to extermination after hearing others' replies*".[285] Moreover, the study also suggested that some of the members of the group were not initially in favor of the neo-Nazi approach, and only shifted towards this extreme view after encountering the forums.[286] In these kinds of online spaces, hate speech became a method to consolidate hateful notions and create a sense of community.[287] "*Digital hate culture builds on a cultivation of common sense amongst its audiences that ultimately seeks to radicalize those who listen*".[288] The result of this unity of scope and radical thinking creates what is referred to as a "swarm": a community of in like-minded individuals united by a common spirit and, generally, by a conspiratorial world view. For example, the concept of white genocide unites audiences of neo-Nazis and radical right-wing individuals and has been used to provide a framework of thought to white people who firmly believed they were under attack from other groups (other white people who support diversity or non-white people).[289]
Other than the creation of interest groups, the internet also facilitates the so-called mob mentality. In the cyberspace, the conditions accelerate and enhance all dangerous groups behaviors directed at specific categories of people that inhabit the internet. For example, cyber sexual harassment and hate speech attacks are quickly experienced by women who assert their opinions, especially in male-dominated fields, and often include graphic insults and rape threats. Moreover, the attacks are commonly delivered in multiple platforms, "*occur at unusually high levels of intensity and frequency*" and are often carried out for months or years.[290]
Another feature of the Internet is the instantaneousness. Compared to other types of media, the internet allows to have a thought or a feeling and deliver it in form of speech online to large audiences within seconds.[291] This attribute of online speech creates a situation in which the limit between the public and private spheres is almost erased. In fact, as Tom Clucas states in the article "Don't feed the trolls", the rise of social media has acted as an encouraging factor in allowing people to project

---

[285] Keats Citron, D. 2014, *Hate Crimes in Cyberspace*, Cambridge, London, Harvard University Press, p. 64
[286] Wojciezak, Magdalena. 2010, "'Don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism", New media and society Sage, Vol 12 No. 4, pp. 643
[287] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 302
[288] Ganesh, Bharath. 2018, "The Ungovernability Of Digital Hate Culture." *Journal of International Affairs*, vol. 71, no. 2, pp. 33
[289] Ibid. p. 36
[290] Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union, p. 27
[291] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 304

their private thoughts and feelings into the public sphere.[292] As for the previous characteristics, it is worth noticing that this phenomenon is not necessarily negative. In fact, the immediacy of social media communication can result in a rise in the democratization of the public sphere, for examples in cases in which the opinion expressed by users and the backlash for specific acts or words leads to apologies or shifts in the behavior of public figures. However, the mixing of the private and public sphere in social media can transform "*areas of the public sphere into an unregulated space where unjustified prejudice and legitimate, reasoned opinion become interchangeable*".[293] Moreover, the immediacy of the response encourages the so-called "gut reactions", with a sensibly higher use of insults, threats or prejudiced language towards the targeted group or person.[294]

Brown provides an interesting argument in relation to the above-mentioned characteristics of online hate speech. In his view, the combination of these features of the new media is what makes the difference between offline and online hate speech. "*[T]he world of online communication is special because it combines anonymity, lack of physical presence, being relatively cheap and easy to use, and the capacity for instantaneous publishing*".[295] The risk of being identified on the internet is significantly lower than it is in normal face to face interaction, and the possible immediate consequences of the speech are eliminated by the physical distance between the speaker and the audience, which enhances the person's willingness to engage in spontaneous acts of hate speech. Finally, the accessibility of the internet and the immediacy of the publication provide further conditions that explain the lack of restraint by the speakers.[296]

All the previously mentioned characteristics are not inherently harmful. In fact, social media has had a democratizing effect on society, being a readily available and easy tool. It has been used extensively, and continues to be used, by activists to battle regimes and dictatorships.[297] However, this democratizing effect comes with high risks, especially in relation to hate culture. In the cyberspace, anyone can turn themselves into a "celebrity", amassing audiences and spreading their message, but the lack of control that the internet entails make the possibility of spreading hateful and violent messages almost infinite.[298] Digital hate culture takes advance of the "*interstitial zone in which regulation of content is contested between governments and the private sector*"[299] and provides an

[292] Clucas, Tom. "'Don't Feed the Trolls': Social Media and the Limits of Free Speech." *Violence and Trolling on Social Media: History, Affect, and Effects of Online Vitriol*, edited by Sara Polak and Daniel Trottier, Amsterdam University Press, Amsterdam, 2020, p. 54

[293] Ibid. p. 55

[294] Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 304

[295] Ibid. p. 306

[296] Ibidem

[297] Ganesh, Bharath. 2018, "The Ungovernability Of Digital Hate Culture." *Journal of International Affairs*, vol. 71, no. 2, p. 30

[298] Ibid. p. 31

[299] Ibid. p. 37

easy pathway for people to direct their anger at selected targets, favoring the growth of online hate speech cases and, more in general, of hate crimes.[300]

2.1.2 Multilateral regulation for online hate speech

The nature of the internet, the rate at which content spreads and the transnational reach it has made the application of national jurisdiction extremely challenging. Although all the previously cited conventions can be applied to online hate speech, the singular characteristics of this form of hate speech would require more specific instruments, aimed at regulating the cyberspace. As explained in the previous paragraph, the consequences of online hate speech are often sustained in countries that are not the country where the content was uploaded or created, and the decisions of national courts often cannot be applied to foreign companies. For these reasons, a supranational decision-making system would seem like the most effective response to the problem. However, this collaborative system has been impossible to put in place, mostly because of the firm opposition of the United States. In fact, the US approach to the issue of hate speech is based on a central principle in place since the 1960s: that speech must not be regulated on the grounds that the message conveyed is unacceptable.[301] Although this approach is a minority view within the western democracies, the position of the US in the international system and the fierce opposition to the bans on speech has turned into a generalized issue in regulating this form of hate speech even in the contexts in which other states tried to create an international regulatory framework.[302]

At the UN level there are two resolutions from the Human Rights Council that acknowledge the issue of online hate speech and harassment. The 2016 resolution on the promotion, protection, and enjoyment of human rights on the internet asserts that the recognized rights of people offline must also be protected online.[303] Later, in 2018, the commission adopted resolutions on the same topic in which several forms of cyberviolence were included. The committee also stressed the importance of "*combating advocacy of hatred that constitutes incitement to discrimination or violence on the Internet, including by promoting tolerance and dialogue*".[304]

---

[300] Ibid. p. 42

[301] Heinze, Erik. 2013, "Review Essay - Hate Speech and the normative foundations of regulation", International Jurnal of Law in Context, Vol 9, pp. 591

[302] Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 236

[303] Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union, p. 50

[304] UN Human Rights Council, *Resolution on the promotion, protection and enjoyment of human rights on the Internet*, 1 July 2016, A/HRC/RES/32/13, para. 1.

The first multilateral agreement that tries to counter the phenomenon of cybercrime was the Council of Europe's Convention on Cybercrime, entered into force on 1 July 2004 and ratified by 66 countries as of today. The convention was drafted by the Committee of Experts on Crime in Cyberspace between 1997 and 2000 and was based on previous Council of Europe recommendations on cybercrime. The convention's main objective, as explained in the preamble, is to "pursue a common criminal policy aimed at the protection of society against cybercrime, especially by adopting appropriate legislation and fostering international co-operation".[305] In other words, the main objectives are to create common definitions for crimes committed on the web and define common rules on the investigation of such crimes, with the aim of harmonizing national legislations, and determine the ways in which to tackle the issue of investigation and prosecution through international cooperation.[306]

The first part of the treaty relates to criminal offences and concentrates mainly on the definition of those offences. The convention identifies nine offences divided into four categories. Six of the offences fall into the first category: offences against confidentiality, integrity and availability of data or computer systems.[307] These offences all concern criminal acts in which the target is the computer or data system, as for example the hacking of a computer system. They include all types of illegal access, interception, interference, and misuse of devices. To be considered an offence, these acts must be conducted intentionally and unlawfully, without authorization by state authority. The second category of offences (integrity and availability of data or computer systems) is the computer equivalent of fraud and forgery, two manipulation-based criminal offences that are generally carried out outside of the internet but defined in traditional norms in terms that do not justify their application to the acts perpetrated via computer. This, combined with the awareness that the damages of fraud and forgery online can be suffered by an extremely large amount of people and can result in the manipulation of assets administered by computer systems, ensured that these criminal acts would be defined in the convention.[308] The third category identifies illegal content, and concerns mostly child pornography, identified as the most dangerous type of illegal content in the cyberspace. The convention contains provisions that concern many different aspects of the issue, from possession to distribution to production of the content.[309] Although most states already criminalize this kind of offence extensively, the convention is a useful tool to combat a phenomenon that is increasingly

---

[305] Council of Europe Website: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=185 [Accessed 14 Jaunary 2022]

[306] Council of Europe, *Convention on Cybercrime*, 23 November 2001, available at: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=185

[307] Ibidem

[308] Csonka, Peter. 2006, "The Council of Europe's convention on cybercrime and other European initiatives", *Revue Internationale de droit penal*, vol. 77, pp. 486

[309] Ibidem

occurring on the internet. The fourth and last category is once again tied to content, but contrary to the prior category, in this case the convention seeks to protect it. In fact, this type of offences relates to infringement of copyright and related right, an ever-present issue on the internet, where content is often reproduced without the prior authorization of the author.[310]

The convention initially included an Internet hate speech protocol, which sought to create a multilateral harmonized framework for addressing different types of illegal content other than child pornography, racist propaganda, and holocaust denial in particular. However, to obtain the USA's signature, the hate speech protocol had to be removed from the main text of the convention, which was introduced in a separate protocol, the already mentioned Additional Protocol to the Convention on Cybercrime in which parties are required to criminalize racist and xenophobic carried out through the use of the internet.[311] The framework provided by the convention on cybercrime, in connection with the additional protocol, provides for a quite effective European legislation on the matter. However, the US is not part of the protocol, as the contents of it are viewed as inconsistent with its constitutional guarantees.[312] For this reason, even if the European framework seems strong, the actual application of the articles contained in the convention and in the protocol proves difficult, as many hate sites and hateful contents originate in the US, where the European legislation has no reach. At the same time, American legal jurisprudence is strongly against the use of bans on speech, regardless of the hateful nature it might have. Moreover, the US is unlikely to extradite citizens if the crimes committed are legal in US territory, which makes the application of national laws for contents originated in the US virtually impossible.[313]

Another interesting multilateral initiative regarding online hate speech is the Freedom Online Coalition. Created in 2011 by the Netherlands, the freedom online coalition is a group of governments "*committed to the human rights and fundamental freedoms proclaimed in the Universal Declaration of Human Rights" and to "working together [...] to support Internet freedom and protect human rights online worldwide*"[314] The coalition now counts 34 governments, including many EU states, the USA and some African, Asian and American countries. To enact the commitments, in 2014 the group has adopted a set of recommendations aimed at protecting online freedoms: the Tallin agenda. In the document, the governments commit to working together on a range of different measures to protect online freedoms. Although related to intern, the document does not mention hate speech and only

[310] Ibid. p. 488
[311] Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 237
[312] Ibid. p. 236
[313] Henry, S. Jessica, 2009, "Beyond free speech: novel approaches to hate on the Internet in the United States, *Information and communications Technology law*, vol. 18, no. 2, p. 241
[314] Freedom Online Coalition website: https://freedomonlinecoalition.com/ [Accessed 15 January 2022]

calls upon governments to halt "*illicit filtering, blocking and monitoring of opposition voices and other repressive measures utilized to restrict freedom of expression and organization online in contravention of international human rights obligations*".[315]

This coalition could potentially become a useful multilateral forum for agenda setting on cyber related issues. However, the growth of the coalition has been slow, possibly because of a deliberate decision from the coalition itself to accept only the members that are able to meaningfully contribute to the coalition, and lacks diversity, which makes the agenda setting less legitimate as it may seem like a Eurocentric approach to the issue.[316] Moreover, the strong First Amendment tradition of the USA and the opposite European approach to hate speech makes the possibility of a strong policy position on the matter by the coalition highly unlikely.

2.2.3 The responsibility of private companies

The regulation of hate speech online is, now, heavily reliant on each platform's standard and rules as there is no accordance on how to regulate content between states, with European countries being mostly in favor of a regulation and the US strongly opposing the bans, since "*the prevailing view among media professionals and scholars is that the Free Speech Principle shields even the vilest forms of hate speech*".[317] The already mentioned EU code of conduct signed by the major online companies in 2016 has proven useful to enhance and coordinate the efforts made by private companies to counter the spread of hate speech, but it has not been entirely effective. The central aim of this instrument is to ensure that the IT companies that sign it are committed to countering hate speech and provide explicit rules and guidelines on the matter.[318] The push for the implementation of the code of conduct and of the various community guidelines on social media resulted from the rise of intolerant speech towards refugees after the terrorist attacks of the 2010s. Twitter's terms of service, adopted in 2016, state, regarding hate speech, that any hateful conduct that promotes violence, threats, or direct attacks to other people is prohibited on the platform.[319] Similarly, in 2018, Facebook released the new rules on content allowed on the platform, stating that hate speech would not be allowed because "*it creates an environment of intimidation and exclusion and [...] promote[s]*

---

[315] Ministers of the Freedom Online Coalition, *Reccomendations for Freedom Online*, Tallin, April 28, 2014, art. 3

[316] Ülgen, Sinan. 2016, FREEDOM OF EXPRESSION ONLINE GOVERNING CYBERSPACE: A Road Map for Transatlantic Leadership, *Carnegie Endowment for International Peace*, 2016, pp. 26

[317] Cohen-Almagor, R. 2015, *Confronting the Internet's Dark Side – Moral and social responsibility on the free highway*, New York, Cambridge University Press, p. 165

[318] European Commission website: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en [Accessed 16 January 2022]

[319] Twitter terms of service website: https://twitter.com/en/tos [Accessed 16 January 2022]

real-world violence"[320] while YouTube prohibits hate speech on its website, using a system that relies on user's reports.[321] However, the ever-present issue of the lack of a clear and consistent definition of what hate speech is makes the understanding of the effectiveness of such approaches difficult.[322]
To complicate the situation even further, there is a lack of consensus on which strategies should be adopted by country to achieve the goal of countering hate speech. In enforcing their rules, social media platforms rely generally on three systems: artificial intelligence, user reporting and content moderators.[323] The first system, based on the automated detection of the content, relies mostly on dictionary-based approaches: specific words are used to detect hate speech, usually insults and slurs. However, considering that the use of codewords or other strategies could make this system too subject to failure, most systems include other approaches, such as "bag of words" and "n-gram" which analyze the frequency of words that appear in a corpus of training data set and avoid the misclassification of words that could be used in various context or that could have been misspelled. Although useful, these techniques are not particularly effective in identifying more subtle forms of hate speech, as they rely heavily on the words used. Moreover, most of these techniques have been tested on only one major website, usually Twitter as its structure favors data collection, and on English language contents.[324]
Therefore, although the standards and guidelines are now present in all the major social media platforms, the effectiveness of such regulations is debated. Different studies have proved that the bans in specific websites effectively reduced the volume of hate speech on the platforms.[325] However, when shifting the focus from a single platform to a more global view of the internet, it is possible to observe that the banned hate speech did not actually disappear, it either moved to a less regulated platform or it changed the communication strategy by trying to avoid the content regulations (for example, by using code words).[326] Moreover, the bans on speech have proven to be, in some cases, counterproductive, as the "attack" to a user could motivate others who share the same beliefs to unite against the companies or to increase their support to the cause.[327]

---

[320] Facebook community standards website: https://transparency.fb.com/it-it/policies/community-standards/ [Accessed 16 January 2022]
[321] Youtube community guidelines website:
https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436 [Accessed 16 January 2022]
[322] Siegel, A. 2020, *Online Hate Speech,* in Persily, N. Tucker, J ed. *Social Media and Democracy*, Cambridge, Cambridge University Press, p. 59
[323] Council on foreign relations website: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons [Accessed 13 January 2022]
[324] Siegel, A. 2020, *Online Hate Speech,* in Persily, N. Tucker, J ed. *Social Media and Democracy*, Cambridge, Cambridge University Press, p. 59 - 60
[325] Berger, J. M., Perez, Heather. 2016, "The Islamic State's Diminishing Returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters", *Occasional Paper GW Program on Extremism*, p. 9
[326] Siegel, A. 2020, *Online Hate Speech,* in Persily, N. Tucker, J ed. *Social Media and Democracy*, Cambridge, Cambridge University Press, p. 72
[327] Ibid. p. 73

Another central issue with the phenomenon is the liability for online hate speech. Usually, hate websites are hosted in jurisdictions that are strategically favorable, in that they are generally tolerant of hate speech. This leads to "*regulatory circumvention and attempts to evade legal liability for hateful content*".[328] Moreover, the implementation of policies is not always transparent, as the companies are ultimately private actors, and the removal of hate speech could be seen as private censorship if not backed by specific governmental and judicial justifications.[329] For these reasons, determining liability for online hate speech from a judicial perspective becomes complicated. The sheer number of actors that could potentially be considered liable for the creation, publication, development, hosting of hateful contents makes the identification of liable subjects even harder.[330] The lawsuit filed by Reginaldo Gonzalez against Twitter, Google, and Facebook for the death of his daughter in the Islamic State's attacks of 2015 in Paris opened the debate on the liability of companies. Gonzales alleged that the companies had to be considered liable for having permitted the use of their social media platforms as tools for the spread of extremist propaganda, which would have enabled the carrying out of the terrorist attacks.[331]

Although experts cited in the lawsuit supported the claims of Gonzalez's team, the regulatory framework of the US suggests that companies should not be liable for the communications that take place on the platforms. In fact, the 1996 Communication Decency Act provides for a minimal regulatory approach to the Internet, and states that "*providers are not owners of the content posted on their platforms, are not liable for any objectionable content, and should not be held responsible for any action taken to limit access to that content*".[332] However, another US law (U.S. Code § 2339(A)) creates a liability for persons that facilitate material support to terrorist actors, which raises some questions on the possibility of application of this precedent to cases of liability of internet companies.[333] Similar questions can arise from the EU framework, in which the Code of Conduct notes that public incitement to violence or hatred are punishable offences. In this case, incitement could be intended as the distribution of illegal contents, therefore calling for a certain amount of liability for internet companies. However, the EU E-Commerce Directive provides some clarifications to the European framework on liability of companies, stating that member states may

[328] McGonagle, Tarlach. 2013. "*The Council of Europe against online hate speech: Conundrums and Challenges*". Republic of Serbia, Ministry of Culture and Information, Expert paper for the Council of Europe Conference of Ministers responsible for Media and Information Society, 'Freedom of Expression and Democracy in the Digital Age: Opportunities, Rights, Responsibilities', Belgrade, 7-8 November, p. 27
[329] Ibid. p. 28
[330] Ibidem
[331] United States Court of Appeal for the Ninth Circuit, Mehier Taamneh; Lawrence Taamneh; Sara Taamneh; Dimana Taamneh, V. Twitter, Inc.; Google Llc; Facebook, Inc., No. 18-17192, D.C. No. 3:17-cv-04107-EMC
[332] Softness, Nicole. "Terrorist Communications: Are Facebook, Twitter, and Google Responsible for the Islamic State's Actions?" *Journal of International Affairs*, vol. 70, no. 1, 2016, p. 209
[333] Ibidem

not impose obligations to providers to monitor the information they transmit, nor to seek for illegal activities taking place on their platforms. What this implies is that member states are not allowed to create such types of obligations unless it is to prevent specific types of illegal acts and activities that are foreseen in national law.[334] Further, article 12 of the same directive states that member states should:

> ensure that the service provider is not liable for the information transmitted, on condition that the provider: (a) does not initiate the transmission; (b) does not select the receiver of the transmission; and (c) does not select or modify the information contained in the transmission.[335]

and article 14 asserts that states shall ensure that the provider is not liable for the information it stores, unless they are aware of the illegal activities or information, or they act to remove the information when it becomes aware of its presence.[336] These normative precedents and frameworks seem to suggest that in general, Internet providers should not be considered liable for possible illegal conducts. At the same time, it is undeniable that the effectiveness of the regulation of the cyberspace is highly dependent on both public and private means, which calls for a shared responsibility between public authorities and private companies on the matter.[337]

2.3 Biases in Hate Speech regulation

Sexism has been defined by two European countries, Belgium and France, as:

> any gesture or act [...] that is clearly aimed at expressing contempt towards a person, based on his or her sex, or, for the same reason, to consider that person as inferior or essentially reduced to his or her sexual dimension, resulting in a serious violation of his or her dignity.[338]

and "*as any act related to the sex of a person, the purpose or effect of which is to damage her/his dignity or to create an intimidating, hostile, degrading, humiliating or offensive environment*".[339] In

---

[334] Center for IT and IP Law website: https://www.law.kuleuven.be/citip/blog/to-monitor-or-not-to-monitor-the-uncertain-future-of-article-15-of-the-e-commerce-directive/ [Accessed 17 January 2022]
[335] European Parliament and Council Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1, art. 12
[336] Ibid. art. 14
[337] Softness, Nicole. "Terrorist Communications: Are Facebook, Twitter, and Google Responsible for the Islamic State's Actions?" *Journal of International Affairs*, vol. 70, no. 1, 2016, p. 214
[338] Council of Europe Gender Equality Unit, *Background note on Sexist Hate Speech*, 1 February 2016, p. 3
[339] Ibidem

countries in which the definition is missing, other concepts are applied to criminalize or sanction sexist behavior, such as anti-discrimination, hate motivation for crimes, incitement to hate, or harassment.[340] Hate speech on the basis of sex (or gender) has not been addressed by any international or EU instrument in a specific manner and, as for hate speech in general, there is no consensus on the definition. In fact, the Council of Europe definition does not mention sex or gender in its definition of hate speech, although the definition includes a broad concept of "other forms of hatred based on intolerance" that could be applied in sexist hate speech cases.[341] However, sexist hate speech and hate speech on the basis of sex or gender are prevalent, especially in the online space, but they are often overlooked or only considered when in conjunction with other forms of hate speech, often racist or religious.[342] To understand the extent and the limits of the protection against hate speech under hate speech law, this paragraph will focus first on the protection from discrimination on the basis of sex, and on the main existing international and regional law instruments for the protection against hate speech on the basis of sex/gender. The analysis will then focus on the phenomenon of sexist online hate speech and on the effectiveness of the protection of women and LGBTQ people under hate speech laws.

2.3.1 Protection from discrimination and hate speech on the basis of sex

The right to equality and non-discrimination is recognized in most international human right law instruments. Article 2 of the Universal Declaration of Human Rights states that "*Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status*".[343] Similarly, the ICCPR establishes the principle in articles 2 and 26, creating an obligation for states to ensure the rights in the convention with no distinction of any kind, and stating that the law "*shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground*".[344] These articles are applicable to any instance of discrimination, but do not provide a definition of discrimination. Such definition is contained in other UN or international treaties that relate to the matter, such as the CERD against racial discrimination

---

[340] Ibidem
[341] Ibid. p. 4
[342] Weston-Scheuber, Kylie. 2012, "Gender and the Prohibition of Hate Speech", *Law and Justice Jurnal*, vol. 12, no. 2, pp. 133
[343] UN General Assembly. (1948). *Universal declaration of human rights* (217 [III] A). Paris. Available at: https://www.un.org/en/about-us/universal-declaration-of-human-rights.
[344] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx.

and the Convention on the Elimination of All Forms of Discrimination against Women. In the CERD racial discrimination is defined as:

> any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.[345]

Similarly, the CEDAW defines discrimination against women as:

> any distinction, exclusion or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field.[346]

The CEDAW is the main international convention that has as at its core the protection of women from discrimination. It was adopted in 1979 and entered into force in 1981. The convention was the final step of the work carried on by the United Nations Commission on the Status of Women, established in 1946 with the aim of monitoring the situation of women and promoting their rights.[347] The convention establishes, through its 14 articles, an agenda for equality which covers three dimensions: the legal status of women, reproductive rights, and the recognition of the influence of tradition on the restrictions put upon women. The broadest attention is given to the legal status of women, with the convention restating some of the provisions of previous international conventions (the Convention on the Political Rights of Women and the Convention on the Nationality of Married Women) that guaranteed the right to vote, to hold public office, and to statehood irrespective of marital status. Moreover, the convention highlights women right to non-discrimination in education, employment, social and economic settings.[348] Regarding reproductive rights, the convention calls for "*a proper understanding of maternity as a social function and the recognition of the common responsibility of men and women in the upbringing and development of their children*".[349] Lastly, states parties are

[345] UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, p. 195, available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx.
[346] UN General Assembly, *Convention on the Elimination of All Forms of Discrimination Against Women*, 18 December 1979, United Nations, Treaty Series, vol. 1249, p. 13, available at: https://www.ohchr.org/EN/ProfessionalInterest/Pages/CEDAW.aspx
[347] Ibidem
[348] Ibidem
[349] Ibidem

obliged to work on the change and elimination of cultural stereotypes and patterns "*which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and women*".[350]

The implementation of the CEDAW is monitored by the Committee on the Elimination of Discrimination against Women. The committee is composed by twenty-three independent experts who oversee the implementation by analyzing state reports, individual complaints, interstate complaints and requests. The committee also issues general comments, open letters, and statements.[351] Moreover, with the adoption of the CEDAW optional protocol in 1999, the committee can receive individual communications by women victims of human rights violations, thus strengthening the protections afforded to women.[352]

The CEDAW and the optional protocol are the central international law instruments that relate to the rights of women and their protection; however, neither one mentions hate speech or similar concepts in their texts, therefore the applicable instruments for hate speech on the basis of sex remain those mentioned in chapter 1. Moreover, there is no United Nations convention concerning discrimination based on sexual orientation and gender identity. The United Nations Human Rights Council was the first to adopt a resolution on the matter, expressing "*grave concern at acts of violence and discrimination, in all regions of the world, committed against individuals because of their sexual orientation and gender identity*".[353] In 2013, the Office of the United Nations High Commissioner for Human Rights launched the Free & Equal campaign, advocating for the end of homophobia and transphobia and fair treatment for the LGBTQ community.[354] Although homophobic and transphobic hate speech are prevalent, especially in online settings, there has been no direct mention by the UN so far.

At a regional level it is possible to find more specific instruments for the protection from discrimination on the basis of sex. The Council of Europe addressed the issue of hate speech based on sex in the definition provided by the ECRI in which sex, gender, sexual identity and sexual orientation are included in the list of personal characteristics.[355] The Council of Europe has addressed the importance of the intersectional nature of discrimination in Recommendation 15 by the ECRI

[350] Ibidem

[351] International Justice Resource Center website: https://ijrcenter.org/un-treaty-bodies/committee-on-the-elimination-of-discrimination-against-women/ [Accessed 17 January 2022]

[352] Gomez Isa, Felipe. 2003, "The Optional Protocol For The Convention On The Elimination Of All Forms Of Discrimination Against Women: Strengthening The Protection Mechanisms Of Women's Human Rights", *Arizona Journal of International and Comparative Law*, vol. 20, no. 2, pp. 291 - 321

[353] UN Human Rights Council, Resolution on Human rights, sexual orientation and gender identity, 14 July 2011, A/HRC/RES/17/19

[354] United Nations Free and Equal website: https://www.unfe.org/about-2/ [Accessed 17 January 2022]

[355] Council of Europe, European Commission against Racism and Intolerance (ECRI) General Policy Recommendation No. 15 on combating hate speech adopted on 8 december 2015.

stating "*Conscious of the particular problem and gravity of hate speech targeting women both on account of their sex, gender and/or gender identity and when this is coupled with one or more of their other characteristics*".[356]

More recently, in 2019, the Committee of Ministers issued a recommendation on preventing and combating sexism in which they noted that "*discrimination on the grounds of sex and/or gender constitutes a violation of human rights and an impediment to the enjoyment of human rights and fundamental freedoms*" and that "*sexism is a manifestation of historically unequal power relations between women and men […] widespread and prevalent in all sectors and all societies*".[357] The committee then encourages governments of member states to take measures to combat and prevent sexism, encouraging stakeholders to implement the appropriate legislation.[358] In issuing this recommendation, the committee states that sexist hate speech plays a role in escalating or inciting threatening acts, that could result in sexual abuse or violence, which makes the creation of norms and standards an obligation under international human rights law.[359] Furthermore, the committee refers to international and regional instrument, such as the Council of Europe Convention on preventing and combating violence against women and domestic violence and the United nations convention on the elimination of all forms of discrimination, in which "*[t]he need to tackle sexism, sexist norms and behaviour and sexist speech is implicit*".[360] Despite this, the recommendation is clearly directed at condemning hate speech against women (on the basis of sex) and girls, while the dimension of gender and sexual orientation is not directly mentioned.

Lastly, it is important to mention the new Gender Equality Strategy by the Council of Europe. This strategy was adopted in 2018 and has six strategic objectives:

1) Prevent and combat gender stereotypes and sexism
2) Prevent and combat violence against women and domestic violence
3) Ensure the equal access of women to justice
4) Achieve a balanced participation of women and men in political and public decision-making
5) Protect the rights of migrant, refugee and asylum-seeking women and girls
6) Achieve gender mainstreaming in all policies and measures[361]

[356] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union

[357] Council of Europe Committee of Ministers, Recommendation CM/Rec(2019)1 of the Committee of Ministers to member States on preventing and combating sexism (*Adopted by the Committee of Ministers on 27 March 2019 at the 1342nd meeting of the Ministers' Deputies)*

[358] Ibidem

[359] Ibid. p. 4

[360] Ibidem

[361] Council of Europe, March 2018, *Gender equality strategy 2018-2023,* (Adopted by the Committee of Ministers of the council of Europe), p. 15

This strategy builds from the "*legal and policy acquis of the Council of Europe*"[362] as well as from the results of the previous strategy and will be carried out until 2023. The previous strategy had encountered several challenges in its implementation, sexism and sexist hate speech being among them. Thus, the first objective that aims at preventing and combating stereotypes and sexism includes sexist hate speech online and offline, as this phenomenon is both increased by gender stereotypes and central in their maintaining and reinforcement.[363] Hence, the council of Europe will seek to "*continue to address sexist hate speech as a form of sexism, analyze and monitor its impact, in co-operation with other relevant sectors of the Council of Europe*" and "*prepare a draft recommendation to prevent and combat sexism, including guidelines to prevent and combat it online and offline […] addressing sexist language, sexist hate speech, sexism in media and in advertising*".[364]
At the EU level, the European commission published a gender and equality strategy for the years 2020 to 2025, aimed at achieving gender equality in Europe by surpassing gender-based violence, sex discrimination and structural inequality. The strategy reaffirms the centrality of the value of gender equality in the EU, seen as an essential condition for an "*innovative, competitive and thriving European economy*".[365] The document than asserts that the EU is a global leader in gender equality, but even so no Member State has achieved full equality so far, thus it is considered imperative to give impetus to the process through aimed initiatives. In fact, the strategy recognizes that the principle of equality is continuously violated through sexist hate speech and the enforcement of gender stereotypes.[366] The strategy is set to follow a dual approach: targeted measures for achieving equality and a strengthening of gender mainstreaming achieved through the inclusion of "*a gender perspective in all stages of policy design in all EU policy areas, internal and external*".[367] Moreover, in addressing gender-based violence in the strategy, the Commission affirms that they will propose a Digital Services Act based on the same EU Internet Forum that led to the adoption of the EU Code of Conduct, in order to address the responsibility of platforms regarding hateful user-uploaded content that has effects on the daily lives of women and girls. The act would become a horizontal instrument aimed at creating a framework of layered responsibilities for the various actors present in the online domain.[368] To achieve this goal, the European parliament already voted three resolutions in 2020 calling for a revision of the E-Commerce directive, the introduction of content management rules and the update of the general legal framework on digital services. Moreover, the appointed Rapporteur

---

[362] Ibid. p. 9
[363] Ibid. p. 16
[364] Ibid. p. 19
[365] European Commission, March 2020, *A Union of Equality: Gender Equality Strategy 2020-2025*, Brussels
[366] Ibid. p. 2
[367] Ibidem
[368] European Parliament website: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act [Accessed 18 January 2022]

for the Digital Services Act Christel Schaldemose affirmed the need to introduce stricter rules on the online environment. However, so far, the focus has been put mostly on the commercial aspect of online services, with the rapporteur suggesting a ban on targeted advertisement.[369] Nonetheless, the act could become a central tool in the fight to online hate speech. The act is now awaiting the first reading in the parliament, announced for the 8th of February 2022.[370]

The European parliament also addressed the issue of homophobic hate speech in the already mentioned resolution on public discrimination and hate speech against LGBTI people, expressing concern and calling for a commitment to diminishing discrimination and monitoring the situation. This recommendation does not have a tangible outcome, it does however provide an understanding of the general approach of the EU parliament to the issue.[371]

Finally, moving from the European framework, it is worth noticing that the Interamerican Commission has addressed the issue of hate speech and incitement to violence against the LGBTQ community in chapter IV of the report "Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas" of 2015. In the report the Commission notes that the Convention of Belém do Para is the only InterAmerican instrument that provides a definition of violence against a group, women, defined as "*any act or conduct based on gender, which causes death or physical, sexual or psychological harm or suffering to women*".[372] The commission reiterates that this type of violence is connected to patterns of socially entrenched inequality, which are pervasive also for non-normative orientations and identities. The Commission then refers to a report by the UN Special Rapporteur on Violence against Women which affirmed that, in general, "*societal beliefs that claim that one group of people is superior to another group can be a form of structural violence*",[373] whether it be referring to structural violence against women or against the LGBTQ community. Thus, the IACHR notes that, even though sexual orientation and gender identity are not included in the convention of Belém do Pará, the convention is to be considered a living instrument. Therefore, the commission considers:

---

[369] European Parliament website: https://www.europarl.europa.eu/news/it/press-room/20211210IPR19209/digital-services-act-safer-online-space-for-users-stricter-rules-for-platforms [Accessed 18 January 2022]

[370] European Parliament website: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2020/0361(COD)&l=en [Accessed 18 January 2022]

[371] European Parliament resolution, 2019/2933(RSP) of 18 December 2019 on public discrimination and hate speech against LGBTI people, including LGBTI free zones, P9_TA(2019)0101.

[372] OAS, IACHR, *Violence Against LGBTI Persons*, Official records; OEA/Ser.L, 2015, p. 45

[373] UNGA, *Report of the Special Rapporteur on violence against women, its causes and consequences*, A/HRC/17/26 (2 May 2011), para. 28

> that when Article 9 of the Convention of Belém do Pará speaks of the State obligation to take special account of factors of special vulnerability to violence, listing certain examples "among others," these factors would necessarily include sexual orientation and gender identity.[374]

Moreover, the report acknowledges the growing issue of hate speech against the LGBTQ community and provides an overview of the InterAmerican legal standards on hate speech, in order to establish a basis for the understanding of the concept and the creation of effective responses.[375]

3.2.2 Sexist hate speech online

Digital technologies have had an impact in many areas of people's lives, making changes in the way individuals express themselves and in how they access or share information. These changes could be an invaluable tool in advancing equality, especially for members of specific discriminated groups. However, along with the positive effect on the empowerment of minorities, the internet has facilitated the expression of hateful messages, insults, and other forms of violence against women and other sexual minorities.[376] There is a significant difficulty in aggregating and comparing data regarding cyber violence against women, due to the uneven approaches to the criminalization of this type of violence.[377] The monitoring of the implementation of the Code of Conduct on countering illegal hate speech online in the EU of 2018 underlines how hate speech on the basis of gender amounts to 3.1% of reports on social media platforms. This percentage includes instances of hate speech towards women that are targeted for their gender as well as for other characteristics (sexual orientation, nationality, background) and refers only to reported cases.[378] According to other surveys, 11% of women in the EU have experienced cyber violence and harassment starting from the age of 15 and 4% have experienced cyberstalking.[379] Although data is scarce, it is possible to say that women are much more likely to experience severe types of cyber violence and harassment compared to men, as well as more likely to suffer more traumatic consequences from the violence.[380] Moreover it is important to say that women do suffer disproportionately from cyberviolence and hate speech, and

[374] OAS, IACHR, *Violence Against LGBTI Persons*, Official records; OEA/Ser.L, 2015, p. 46
[375] Ibid. p. 128
[376] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union
[377] Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union, p. 40
[378] Ibid. p. 42
[379] Ibidem
[380] Ibid. p. 1

that the violence is often coupled with other type of violence: sexuality, disability and age based harassment.[381]

Cyber-violence and hate speech, as any other form of violence against women, have short-term, long-term, and intergenerational effects. Regarding the women's health and social situation:

> Amnesty International found that of the women who experienced abuse or harassment online, 41% of responding women felt that their physical safety was threatened [and] 1 in 2 women experienced lower self-esteem or loss of self-confidence as well as stress, anxiety or panic attacks as a result of cyber violence and hate speech online.[382]

In addition, the online violence and hate speech can result also in damages to the economic situation of women, as the long-term effects can affect the reputation of the victims and damage their quality of life. Furthermore, the constant online harassment can push women to live social media and the internet, even those women who rely on it for their income, which will damage both their economic life and the society in general.[383]

To have a better understanding of what cyber-sexism is, it is necessary to understand the main concepts involved in the definition of this specific type of gender-based discrimination. First, sexism is defined by Bailey Poland as "*a combination of prejudice against persons based on their gender, combined with the privilege and power required to cause harm*".[384] Today's men, considered as a social group, hold most privileges and power (financial and political for example) and this situation allows for their prejudices to be much more likely to hurt, limit and create difficulties for women.[385]

Privilege exists when one group, with specific identifying characteristics, has something that others are denied because of their belonging to a certain identifiable group and not because of "*anything they have done or failed to do*".[386] In other terms, privilege is a set of social advantages that is related to the affiliation to a certain identity group. The overrepresentation of the relation between the group and the social advantages results in a difficulty in noticing privilege, especially by those who carry it, as the association is considered default.[387] In the case of male privilege, the association relates to a greater representation in media and business, as well as easier access to power position and employment in general. Sexism connects to privilege and patriarchal ideology, as it is a set of attitudes that rests on the employment of the assumptions, beliefs, stereotypes, and cultural narratives that are

---

[381] Ibid. p. 30

[382] Ibid. p. 33

[383] Ibid. p. 34

[384] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 2

[385] Ibidem

[386] Johnson, A. 2001, *Privilege, Powe and Difference*, New York, McGraw-Hill, p. 21

[387] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 2

advanced by those situations.[388] Sexism is therefore "*the ability to cause harm to a group while conferring benefits to another group*"[389] by applying the association between the group and the characteristics it is socially associated to. Lastly, cyber-sexism is defined as the application of these concepts to the online space using technology.[390] At the European level, sexist hate speech is defined in the Additional Protocol to the Convention on Cybercrime as "*expressions which spread, incite, promote or justify hatred based on sex*" and the Cybercrime Convention Committee highlighted how cybercrime can entail types of harm that are not found outside of the cyberspace, as the consequences of cybercrime often persist for a long time even after the commission of the crime.[391]

In legal terms, cyber-violence and hate speech online against women are considered a form of gender-based violence, along with other forms of violence such as harassment, stalking, non-consensual image-abuse, and sexist hate speech.[392] At the UN level, the most relevant definition of gender-based violence comes from the CEDAW committee, which defines it as "*violence that is directed against a woman because she is a woman or that affects women disproportionately. It includes acts that inflict physical, mental or sexual harm or suffering, threats of such acts, coercion and other deprivations of liberty*" in General Recommendation 19 and extends the definition stating that gender-based violence manifests in "*a continuum of multiple, interrelated and recurring forms, in a range of settings, from private to public, including technology-mediated settings*" in General Recommendation 35.[393] Concerning specific terminology, the Special Rapporteur on violence against women asserted that it is not yet univocal, therefore the terms "ICT[394]-facilitated violence against women", "online violence against women", "cyberviolence" and "technology-facilitated violence" are all effective terms to indicate the "*gender-based violence against women that is committed, assisted or aggravated in part or fully by the use of ICT*".[395]

At the European level there is no commonly accepted definition for online hate speech against women or cyberviolence, there is however an implicit reference to the concept in some Council of Europe Conventions. The Istanbul convention, the first legally binding agreement that aims at reducing violence against women and partner violence, contains articles that can be applied to cyberviolence and hate speech. Violence against women is defined as:

---

[388] Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield, p. 41

[389] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 3

[390] Ibid. p. 4

[391] Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union, p. 13

[392] Ibid. p. 11

[393] Ibidem

[394] Information and Communication Technology

[395] Ibid. p, 12

> A violation of human rights and a form of discrimination against women and shall mean all acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological or economic harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life.[396]

Intimate partner violence is defined as:

> all acts of physical, sexual, psychological or economic violence that occur within the family or intimate partner unit or between former or current spouses or partners, whether or not the perpetrator shares or has shared the same residence with the victim.[397]

Both definitions can be interpreted to include cyber-violence and gender-based violence against women. Additionally, the convention criminalizes psychological violence, stalking and sexual harassment in a way that allows for the application of the norms to the cyberspace.[398]
As we mentioned, the singular characteristics of social medias and online content fruition and diffusion create a situation in which opinions and ideas can be shared easily in virtually no time by users. Aided by the characteristics of the internet (anonymity for example), cyberviolence and hate speech against women draw from the continuum of violence against women and from the entrenched gender inequalities found in modern societies, especially in the tech sector, which reinforce biases and normalize violence.[399] A large part of the modern online exchanges take place in comment sections, in which communities of users have the possibility to decide what speech is acceptable what needs to be reported and how much offensiveness can be accepted in the comment section, although slightly different logics are applied depending on the platform.[400] It is in these setting that a large part of hate speech directed to women is delivered. In "Hate Speech Against Women Online" by Louise Richardson-Self, the author reported the results of her research on online hate speech against women to prove the connection between the phenomenon and the enforcement and enactment of societal power dynamics.[401] The research centers on comments responding to posts made by The Australian,

---

[396] Council of Europe, *Council of Europe Convention on preventing and combating violence against women and domestic violence*, 11 May 2011, available at: https://www.coe.int/en/web/istanbul-convention/about-the-convention, art. 3
[397] Ibidem
[398] Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union, p. 13
[399] Ibid. p. 20
[400] Clucas, Tom. "'Don't Feed the Trolls': Social Media and the Limits of Free Speech." *Violence and Trolling on Social Media: History, Affect, and Effects of Online Vitriol*, edited by Sara Polak and Daniel Trottier, Amsterdam University Press, Amsterdam, 2020, p. 48
[401] Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield, p. 82

a daily newspaper, Facebook page. First, the author provides the rationale behind the choice of website and page, stating that Facebook was chosen as it is the largest social media platform worldwide, while the choice of the Australian newspaper stems from the author's own familiarity with the context. The author collected the data from September 2018 to February 2019 from articles that were either written about or by a woman, featured the image of a woman or centered on gender issues. Richardson-Self decided to collect data manually, to ensure that the collection would have also captured "*instances of hate speech that were not obviously vituperative*".[402]
In carrying out the research, Richardson-Self found 2881 comments that met the set standards, expressing hate speech towards women with an average of sixteen hateful comments per day. The author then determined, based on the usernames, that more than 70 per cent of comments were left by users with masculine usernames, which meant that "*for every one hate speech comment left by a woman user, over three (3.48) hate speech comments were made by men*".[403] The author then provides an analysis of the contents and rationales of hate speech comments directed at women, identifying nine different reoccurring themes in the analyzed data. The first two groups refer to women in connection to sex. Women are commonly portrayed in the comments as sexual objects rather than people, in both "complementary" and antagonistic ways. In such examples, it seems like the central goal of the comment is to fix the women in a specific hierarchical role in relation to men, "*the object of a subject's desire*".[404] Consequently, in the antagonistic forms of this type of comments, the women are portrayed as undesirable, implying they lack any kind of value. Similarly, there seems to be a tendency to oversexualize women of all sorts of settings and backgrounds, even in instances in which the original post does not provide any perception of the women's moral character.[405] Of a similar nature is the group of comments that reduces women to their bodies. In this case, hate speech instances include "*(a) allusions to female primary and secondary reproductive organs […]; (b) the reduction of women to or inordinate focus on mere body parts […] and (c) their reduction to bodily capacities*".[406] The next two groups unite all those comments in which women are cast as beasts, "*in particular beasts with a sedimented history of application such that the terms now can be considered gendered slurs*" (such as the word "cow"), or as monstrous mythical creatures (such as witches or harpies). [407] The author finds that this sort of comments tends to be directed to posts in which the mentioned women are conveying a message or stating an opinion; therefore, the comment seems to

[402] Ibid. p. 83
[403] Ibid. p. 84
[404] Ibid. p. 86
[405] Ibid. p. 90
[406] Ibid. p. 91
[407] Ibid. p. 93 - 94

be used as a silencing tool, degrading the speaker and thus making her arguments worthless as they come from such an unrespectable source.[408]

The following two categories of comments are defined by the author as the abjectification of women and the concept of failed women. By abjection the author refers to "*that feeling of sickness caused by the proximity of an object that is already designated as disgusting, a sickness that may involve gagging or pulling away*".[409] Thus, the comments of this type could be identified by the presence of the vomiting or nausea emoji, usually directed at famous women, combined often with an explanatory text of the abjection felt by the user. Moreover, "*[t]o abjectify women, commentators would also make comments about cisgender women's aged appearance, their body shape […] and other accusations of gross ugliness or decay*",[410] which suggests that "*women are consistently valued by their attractiveness to men, unattractiveness is thought to be the highest form of insult*".[411] The concept of failed women is instead used mostly referring to lesbian women, gender non-conforming individuals, or more in general queer women. "*Though it is achieved through different figurations, every tactic images Women as somehow lacking in the domain of sexual orientation, gender status and gender expression, and therefore Failed*".[412] Another common instance of hate speech against women online sees the use of cultural stereotypes to discredit the credibility or, more in general, take down women. For example, the use of jokes such as "Will someone please think of the sandwiches" or other similar remarks related to stereotypes are common in comment sections. In this case the author recognizes that the comments "*sit at the limits of the category 'hate speech'*",[413] but asserts that the continuous use of such stereotypes expressed in the present social context contributes to reaffirming such stigmatized images. Finally, Richardson-Self underlines that a common theme in the content is the blatant antifeminism that many hate speakers express in the comments. Feminists are portrayed as coming for men, having an agenda to downgrade men and obtain domination.[414]

Although limited in time and data, this research allows to understand in part what women are subject to when inhabiting online spaces. Cybersexism and online hate speech directed at women tends to create, enforce, and normalize male dominance. This type of abusing behavior is "*often framed as an evanescent phenomenon*"[415] and reduced to a simple prank, not to be taken seriously. However, for the women that are subject to this type of harassment, the experience entails a constant effort to

[408] Ibid. p. 94
[409] Ahmed, S. in Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield, p. 96
[410] Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield, p. 97
[411] Ibid. p. 98
[412] Ibid. p. 99
[413] Ibid. p. 103
[414] Ibid. p. 104
[415] Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books, p. 89

monitor their online presence, as well as a societal pressure to "not take it seriously".[416] In fact, the internet started as a free and scarcely inhabited place where the majority of users were under the assumption that their opinion would not be seen or heard by other people. However, in its present form, the internet is "*as real and as important for most people as offline life*"[417] and the effects of online and offline abuse become increasingly similar. Moreover, the permanence of online contents makes the damages of abusive comments and online sexism even more psychologically damaging, as there is no getting away from it.[418] The constant exposure to this abuse even in contexts that are, theoretically, far from extremist views or hateful websites (the Facebook page of a newspaper for example) results in a constant stress for women that inhabit online spaces.[419] Sexist hate speech "*reinforc[es] and perpetuat[es] hierarchies of identity-based oppression via systematically violent expression constituting a hostile environment*",[420] creating a threat to social peace by accumulating in the environment and resulting in a toxic effect.

2.3.3 The protection of women and the LGBTQ community under hate speech laws: theory vs practice

The lack of standards and explicit protections for hate speech on the basis of gender and sex is reflected on the judicial application of the norms countering hate speech. "*[E]xperts report a very low number of judicial cases of hate speech in general, and specifically in regard to sexist hate speech*".[421] In general, at the international level it is accepted that gender equality includes a formal aspect, achieved through the equal and neutral treatment of people in law, and a substantive aspect, which requires that the application of such laws effectively reduces the disadvantages of either gender.[422] However, the number of cases at the international and national level does not reflect the incidence of the phenomenon of hate speech on the basis of sex.

At the European level, the European Court of Human Rights has judged 60 hate speech cases between 1979 and 2020. Among these, only three were brought to the court by the victims of the hateful speech, while the rest was brought by the utterers and in most cases (62%) the applicants lost the

---

[416] Ibid. p. 90
[417] Ibidem
[418] Ibid. p. 91
[419] Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield, p. 107
[420] Ibidem
[421] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union, p. 157
[422] Edwards, A. 2011, *Violence Against Women Under International Human Rights Law*, Cambridge, Cambridge University Press, p. 158

case, with the court confirming the illegitimacy of the speech found by national courts.[423] The majority of cases judged by the court dealt with violence, religious hatred, ethnic hatred and genocide denial and, considering that some cases have more than one theme, these four categories can be found in 51 of the 60 cases. The theme of terrorism can be identified in 10 cases, totalitarianism in 7, antisemitism in 3 and blasphemy, defamation, and holocaust denial in 5.[424] Only in 3 of the 60 cases judged before 2020 the court dealt with homophobic or transphobic hate speech: Vejdeland and others v. Sweden, Beizaras and Levickas v. Lithuania and Lilliendhal v. Iceland.[425] The first was already discussed in the paragraph regarding the case law of the ECtHR. The second and third case deal with two different aspects of online homophobic hate speech.

Beizaras and Levickas v. Lithuania is one of the very few cases which was brought to the court by the victims. In this case, two men in a relationship alleged that they had been discriminated by the authorities because of the authority's refusal to launch an investigation on the hate comments that flooded the Facebook page of the two men after they had posted a picture of them kissing. The comments called for the castration, killing, extermination and burning of the two applicants. The prosecutor considered the comments "merely expressing their opinion", thus not being liable for prosecution.[426] The domestic courts endorsed the decision of the prosecutor and, moreover, held that "*the applicants' behaviour had been 'eccentric' and deliberately provocative" as they should have realized that the posting of the picture would not contribute to social cohesion in a country in which 'traditional family values were very much appreciated'*".[427] In this case the court held that there had been a violation of article 14 on the prohibition of discrimination, in conjunction with article 8 on the right to respect for private and family life, and of article 13 on the right to an effective remedy. The court found that the two applicants had been discriminated against because of their sexual orientation by the government, since the government did not provide any legitimate justification for the lack of an investigation.[428] The court noted that the sexual orientation of the applicants had played a central role in the lack of an investigation by the government, and that the emphasis put on the "eccentric behaviour" of the applicants by the national court was concerning. The court therefore found a violation of article 13 because of the denial of effective domestic remedies.[429]

The case of Lilliendhal v. Iceland also relates to online hate speech, but was brought to the court by the utterer, an Icelandic national fined by the national courts for the homophobic comments he made

---

[423] The Future of Free Speech website: https://futurefreespeech.com/hate-speech-case-database/ [Accessed 20 January 2022]

[424] Ibidem

[425] European Court of Human Rights, *Factsheet - Hate speech*, Press Unit, June 2020, p. 11 - 12

[426] Beizaras And Levickas V. Lithuania, App. No. 41288/15 (ECtHR, 14 January 2020), para. 18

[427] Ibidem

[428] Ibid. para 152

[429] Ibid. para. 156

in response to an article online regarding the decision of strengthening the school education on LGBT matters. The comments expressed disgust on the matter using derogatory terms for LGBT people. In this case, the applicant alleged that the decision of the national courts breached his right to freedom of expression.[430] The court found the complaint manifestly ill-founded, as the interference with the right to freedom of expression had been prescribed by law and had the legitimate goal of protecting the rights of others. Moreover, the decision of the national court was considered proportional to the nature and severity of the comments in both the severity (a fine) and the amount (approximately 800 euros).[431]

These two cases can provide an insight on the general approach of the court to homophobic hate speech cases and create precedents at the European level for this kind of hate speech; however, the limited number of judged cases compared to the statistically high amount instances in which homophobic hate speech is found, especially in the online setting, highlights an issue in the application of hate speech laws to these instances. Moreover, it is important to notice that none of the cases judged by the court deals with sexist hate speech, nor is it possible to identify a sexist aspect in any of the cases.

To further analyze this aspect of hate speech regulation it is necessary to concentrate on national cases as well. The explicit criminalization of hate speech on the basis of sex or gender can be found only in 14 European states, while that of sexual orientation/gender identity/sex reassignment in 23 states.[432] The incorporation of sex or gender in the prohibitions follows two main approaches: "*explicit incorporation in the enumeration of the grounds included in the definition of the offence; incorporation of sex/gender under the broader notion of 'other group' or similar, through case law or other policy documents*".[433] The prosecution is generally ex officio, as required in the Framework Decision on combating certain forms and expression of racism and xenophobia by means of criminal law.[434] However, national experts report an extremely low number of hate speech cases in national courts, and even less cases regarding sexist hate speech specifically.[435]

For example, in a report on hate crime and hate speech by the Prism project of the EU, the country study regarding France affirms that in general there are not many prosecutions of hate speech, which suggests that the anti-discrimination and hate speech provision are rarely applied by judges.[436] The

---

[430] Lilliendhal v. Iceland, App. No. 29297/18 (ECtHR, 12 May 2020) para. 23
[431] Ibid. para. 48
[432] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union, p. 152 - 153
[433] Ibid. p. 153
[434] Ibid. p. 155
[435] Ibid. p. 157
[436] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 123

country study regarding Spain highlights the absence of a definition of hate crime or hate speech in the Spanish legal system; hate speech cases are usually processed through discrimination laws in cases in which the expression "*represents a form of discrimination or those types of acts that different law provisions penalize*".[437] Moreover, the report highlights how there seems to be a significant lack of awareness and training regarding anti-discrimination legislation and, consequently, hate speech, which explains the limited number of related cases processed.[438] Of a similar nature are the issues reported by national experts in Germany, where it was underlined that the main issues relate to the access to justice, the mobilization of the existing provisions and the application of the same provisions, as the application of the criminal law on the matter is often overlooked in favor of civil law approaches.[439]

Lastly, it is important to note that there are very few states that have laws regulating online hate speech or the online aspects of crimes in general,[440] which reflects on the comparatively low incidence of hate speech on the basis of sex cases that have been judged, as the online dimension of sexist hate speech is extremely significant.

[437] Ibid. p. 243
[438] Ibid. p. 245
[439] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union, p. 156
[440] Ibid. p. 157

Chapter 3
Different national approaches to the regulation of Hate Speech

In this chapter the study will focus on three cases of national regulation of hate speech in which several aspects of the issues analyzed in the previous chapter are going to be highlighted. The analysis will first center on the German legal framework for hate speech, one of the most complete in Europe. It will later move on to France's framework of laws about hate speech and the "Avia Bill" on combating hateful online content, adopted in May 2020, that was then quickly declared unconstitutional by the French Constitutional Council. The next paragraph will focus on the Italian case, analyzing the provisions that can be applied to hate speech cases in the country, and will outline a draft law on the protection of LGBTQ people, DDL Zan, and its complicated path in the parliament. The last paragraph will center on an in-depth analysis of the American first amendment on freedom of speech, on the jurisprudence of the supreme court on hate speech and first amendment cases and on a landmark case on freedom of speech, Snyder v. Phelps.

3.1 Germany: the legal framework on hate speech and the NetzDG

Germany has ratified all the UN conventions cited in previous paragraphs[441] and is a member state of the European Union, thus bound by the analyzed EU instruments. Furthermore, the German government ratified both the Council of Europe Convention on Cybercrime and the additional protocol to the convention.[442] From an international standpoint, it can be said that Germany is on par with the major international instruments for combating hate speech.
The definition of the term hate speech (Hassrede in German) in the country is still open, as there is no legal definition of the concept. In a *Bundestag* document of 2018 hate speech and incitement to hatred have been defined, based on Recommendation (97)20 of the Council of Europe on hate speech, including:

> the intentional disparagement and threats – in word, image and sound – against certain people or groups of people due to their affiliation to a minority, as well as all expressions of hatred, which

[441] United Nations Human Rights Office of the High Commissioner website: https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=62&Lang=EN [Accessed 30 January 2022]
[442] Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=189 [Accessed 30 January 2022]

are based on intolerance, which propagate or incite hatred, promote hatred or any justification of hatred.[443]

The definition does not require violence or threat of violence for the expression to qualify as hate speech and even if there is no legal definition of the term, there are several provisions that can be applied to cases of hate speech without requiring violence as a characterizing part of the act.[444]
This paragraph will analyze the general legal framework of Germany on hate speech, with a specific focus on a highly debated provision, the NetzDG, entered into force in 2018 concerning online hate speech and the responsibility of private companies.

3.1.1 The German legal framework for hate speech

"*In Germany, the fundamental right of freedom of expression is not unlimited. It finds its limits (already) when human dignity is attacked. The expressions of opinion subsumed under the term "hate speech" may therefore well constitute criminal offences*".[445] In fact, Germany has one of the strictest set of laws for speech, as a response to the Nazi experience of the last century.[446] The German Constitution provides, along with the right to freedom of expression, some specific limitations to the right. In article 5, relating to freedom of expression, paragraph 5(1) states: "*these rights shall find their limits in the provisions of general laws, in provisions for the protection of young persons, and in the right to personal honor*".[447] To understand which kinds of expression are awarded the protection under the German Constitution, it is necessary to analyze the Constitutional Court ruling on a case of Holocaust Denial.[448] In this case, the German high Court stated that "*the Federal Constitutional Court has consistently ruled, therefore, that protection of freedom of expression does not encompass a factual assertion that the utterer knows is, or that has been proven to be, untrue*".[449]
The German Federal Criminal Code prohibits public incitement to criminal behavior under Section 111. However, this provision is rarely applied by courts, as they ruled that "*simply giving information,*

[443] Hogan Lovells, *The Global Regulation of Online Hate: A Survey of Applicable Laws – Special Report*, Hogan Lovells for PeaceTech Lab (December 2020), p. 123
[444] Ibid p. 124
[445] Ibidem
[446] Politico website: https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/ [Accessed 4 February 2022]
[447] Brugger, Winfried. 2003, "The treatment of Hate Speech in German Constitutional Law (part 1)", *German Law Jurnal*, Vol 4. No. 1, p. 3
[448] BVerfGE 90, 241, 247, Decision of 13 April 1994, Auschwitz Lie Case (Holocaust Denial Case) = Decisions 620, at 625.
[449] Brugger, Winfried. 2003, "The treatment of Hate Speech in German Constitutional Law (part 1)", *German Law Jurnal*, Vol 4. No. 1, p. 13

a statement of political dissatisfaction or engaging in provocation is not sufficient for Section 111”,[450] nor is the endorsement of criminal acts. Similarly, section 140(2) of the same code criminalizes the public approval of certain criminal acts already committed and has been applied by courts, for example in a 2016 case relating to the public approval of the killing of a journalist by ISIS.[451] However, the central provision relating to hate speech in the German Criminal Code is Section 130, on Incitement of Masses. This section criminalizes:

> Whosoever, in a manner capable of disturbing the public peace 1. incites hatred against segments of the population or calls for violent or arbitrary measures against them; or 2. assaults the human dignity of others by insulting, maliciously maligning, or defaming segments of the population.[452]

Further, it provides for the punishability of:

> Whosoever […] (a) disseminates such written materials; (b) publicly displays, posts, presents, or otherwise makes them accessible; (c) offers, supplies or makes them accessible to a person under eighteen years; or (d) produces, obtains, supplies, stocks, offers, announces, commends, undertakes to import or export them.[453]

Thus, the includes incitement to hatred, violence, arbitrary measures and massive abuses directed at a national, racial, religious group, at a part of the population or at an individual and “*if the activity is carried out in a way that is likely to disturb the public peace*”.[454] It should be noted that violence is not a prerequisite for the application of this section of the Criminal Code, and that “*the concretely affected person does not have to feel insulted, and does not have to have any interest in prosecution. It is sufficient for someone to hear the insult and then report it to the police*”. Moreover, the provision expressly requires a liability for the dissemination of materials by posting them or making them accessible, therefore making it applicable for cases of online hate speech.

[450] Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States*, p. 22
[451] Ibidem
[452] Sherloc United Nations Office on Drugs and Crime website: https://sherloc.unodc.org/cld/en/legislation/deu/german_criminal_code/special_part_-_chapter_seven/section_130/section_130.html [Accessed 5 February 2022]
[453] Ibidem
[454] Hogan Lovells, *The Global Regulation of Online Hate: A Survey of Applicable Laws – Special Report*, Hogan Lovells for PeaceTech Lab (December 2020), p. 125

3.1.2 The NetzDG: combating hate speech on social media

The German's Network Enforcement Act (NetzDG in short) is known as a hate speech law that aims at holding social media platforms responsible for the fight to online hate speech. The decision to implement such an act was taken after the publishing of a research carried out by the accredited flagger Jugendschutz.net, who had been assigned the task of assessing online hate crimes by the government in 2016.[455] The accredited flagger investigated content violating the abovementioned Section 130 of the Criminal Code over the course of two periods in 2016 and 2017, and found that "*Facebook removed 39%, YouTube 90% and Twitter 1%*".[456] Taking these findings as a justification, the government of Germany introduced the NetzDG draft law to the Parliament on March 2017.[457] The Act entered into force on the 1 January 2018 and its central purpose is to ensure that 22 statutes already part of the German law are applied to online platforms and the Internet in general, as well as holding IT Companies accountable for their enforcement on their platform.[458] The statutes include the prosecution of insults, libel, slander, public incitement to commit crimes, incitement of masses, violence, defamation, violation of privacy, and other connected categories.[459] Regarding the liability of companies, the act requires Social Networks to "*set up an effective complaint mechanism and to produce a report every six months on how they have handled complaints*".[460] Section 1 of the law defines social networks as "*telemedia service providers which, for profit-making purposes, operate internet platforms which are designed to enable users to share any content with other users or to make such content available to the public*",[461] but excludes journalistic and editorial content providers from the obligations contained in the act.[462] Section 2 provides for the reporting obligations of Social Network Providers and requires them to provide half-yearly reports to be published in the Federeal Gazzette, in order to ensure transparency and facilitate data gathering.[463] In section 3, the act deals with the handling of complaints about alleged unlawful content and states that the provider must "*removes or blocks access to content that is manifestly unlawful within 24 hours of receiving the complaint*" or "*removes or blocks access to all unlawful content immediately, this generally being within 7 days of receiving the complaint*", unless there is reason to believe that the reporting was false

[455] Echikson, W. Knodt, O., 2018, *Germany's NetzDG: A key test for combatting online hate*, CEPS Digital forum, Counter Extremism Project, p. 4
[456] Ibid. p. 5
[457] Ibidem
[458] Tworek, H. Leerssen, P., 2019, *An Analysis of Germany's NetzDG Law*, Transatlantic Working Group, p. 2
[459] Hogan Lovells, *The Global Regulation of Online Hate: A Survey of Applicable Laws – Special Report*, Hogan Lovells for PeaceTech Lab (December 2020), p. 131
[460] Echikson, W. Knodt, O., 2018, *Germany's NetzDG: A key test for combatting online hate*, CEPS Digital forum, Counter Extremism Project, p. 5
[461] German Law Archive website: https://germanlawarchive.iuscomp.org/?p=1245 [Accessed 5 February 2022]
[462] Ibidem
[463] Ibidem

or there needs to be further analysis by a recognized self-regulation institution.[464] "*Failure to meet reporting requirements is an administrative offence under Section 4 of the Act, punishable with fines up to 500,000 Euros*".[465] It should be noted that in article (3) of section 4, the act states that "*The regulatory offence may be sanctioned even if it is not committed in the Federal Republic of Germany*" and thus explicitly provides for the prosecution of companies regardless of their position.[466]

The implementation of the law has been subject to several criticisms by tech industries, activists, and academics.[467] Article 19, an international human rights organization that concentrates on freedom of expression,[468] provided a thorough analysis of the act, expressing its concerns regarding the provision. The first critical point highlighted in the report concerns the definition of Social Network provided in the law. According to Article 19, the definition of Social Networks provided by the law is too ambiguous and "*creates significant uncertainty about who the duties in the Act apply to*".[469] In particular, the organization referred to the words "sharing or making available content", which in theory can be applied to any platform that allows the posting of third-party contents or that connects to third party content, which, according to the organization, results in a lack of clarity of the scope of the provision.[470] Similarly, the organization contested the exclusion of journalistic content from the provision, not because of an unjust treatment, but because of the lack of a definition of what journalistic content is, which once again makes the identification of the scope of the provision difficult.[471]

Another strongly criticized aspect of the provision is that it would encourage the over-removal of content. In fact, it was argued that online platforms, lacking time and expertise, would resort to the removal of content as "*[the costs of the procedure] as well as NetzDG's tight deadlines and heavy fines, platforms would have a strong incentive simply to comply with most complaints, regardless of their actual merits*".[472] Moreover, the act does not provide for any alternative to the solving of disputes other than the blocking or removal of content.[473] Therefore, according to the critics, this kind of approach would result in a clear violation of freedom of speech. For these reasons, in the concluding remarks of its report, Article 19 affirmed that "*the Act, taken overall, [is considered] to*

---

[464] Ibidem
[465] Article 19, 2017, *Germany: The Act to Improve Enforcement of the Law in Social Networks*, Media Against Hate Campaign, p. 22
[466] German Law Archive website: https://germanlawarchive.iuscomp.org/?p=1245 [Accessed 5 February 2022]
[467] Tworek, H. Leerssen, P., 2019, *An Analysis of Germany's NetzDG Law*, Transatlantic Working Group, p. 3
[468] Article 19 website: https://www.article19.org/about-us/ [Accessed 23 January 2022]
[469] Article 19, 2017, *Germany: The Act to Improve Enforcement of the Law in Social Networks*, Media Against Hate Campaign, p. 12
[470] Ibidem
[471] Ibidem
[472] Tworek, H. Leerssen, P., 2019, *An Analysis of Germany's NetzDG Law*, Transatlantic Working Group, p. 3
[473] Article 19, 2017, *Germany: The Act to Improve Enforcement of the Law in Social Networks*, Media Against Hate Campaign, p. 21

be dangerous to the protection of freedom of expression in Germany, and we are particularly concerned that countries with much weaker institutional and legal safeguards for the protection of human rights are looking at this Act as a model for increasing intermediary liability", recommending the repealing of the act.[474]

The concern expressed by the organization has found a partially concrete realization in France where, although the institutional safeguards for human rights are strong, the NetzDG was used as an inspiration for a similar bill on online hate speech, whose development will be analyzed in the following paragraph.

3.2 France: Hate Speech laws and the failure of the "Avia Bill"

France has ratified all the major UN international convention cited in previous paragraphs,[475] and, as a member state of the European Union, is bound by the mentioned EU Instruments. Moreover, the French government ratified both the Council of Europe convention on Cybercrime and the additional protocol to the convention.[476] Thus, from an international point of view, France should have quite a complete legal framework on hate speech, discrimination, and equality.

In France, freedom of expression was considered an inalienable right since the French Revolution and the issuing of the Declaration of the Rights of Man and of the Citizens of 1789. However, as the French government website shows, the French approach to freedom of expression is not that of an absolute nature of the right. In fact, in the French legal framework "*[t]his freedom has limits: racism, anti-Semitism, racial hatred, and justification of terrorism are not opinions. They are offences*".[477]

Regarding hate speech and discrimination, France has a substantive framework of regulation that allows to address the issue with a level of accuracy.[478] However, the current framework of regulation on hate speech and discrimination is not fully able to contrast the phenomena. The effectiveness of the provisions and the jurisprudence on the matters have shown that the application of the laws lacks consistency at the state level and creates conflicts in jurisdiction at the international level, as happened for example in the Yahoo case of 2001.[479]

---

[474] Ibid. p. 24

[475] United Nations Human Rights Office of the High Commissioner website: https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=62&Lang=EN [Accessed 18 January 2022]

[476] Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=189 [Accessed 18 January 2022]

[477] French Government website: https://www.gouvernement.fr/en/everything-you-need-to-know-about-freedom-of-expression-in-france [Accessed 18 January 2022]

[478] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 98

[479] Ibidem

The paragraph will first analyze the French framework of regulation on discrimination and hate speech, and it will then focus on the heavily criticized and ultimately deemed unconstitutional legislation on online hate speech, the Avia Bill.

3.2.1 Overview of the French provisions on discrimination and Hate Speech

In the French Constitution, adopted in 1948 and reformed in 2008, most provisions within the text refers to the structure of the government and of the authoritative bodies of the republic. It does not contain a list of rights and principles; hence, the central source of laws on the matters of discrimination and equality are to be found in the preamble to the constitution of 1958, which recalls previous texts. The principle of equality is recognized in the preamble of the present constitution through a reference to the Declaration of 1789, as well as to the preamble to the constitution of 1946, which prohibits discrimination on the basis of sex, race, belief and trade union activity, by affirming that "*France shall form with its overseas peoples a Union founded upon equal rights and duties, without distinction of race or religion.*".[480] It is important to note that the Constitutional Council has decided that the list of grounds included in the mentioned preamble and, by reference, in the following constitutional text is not to be considered exhaustive and other grounds can be added when considered necessary.[481] Moreover, the current constitution refers to the principle of equality in article 1 by stating that the Republic shall "*ensure the equality of all citizens before the law, without distinction of origin, race or religion*".[482] France also recognizes the superiority of international treaties and conventions over national law in article 55 of the Constitution and establishes the body of the "Defender of Rights" in article 71, which has the role of ensuring that "*the due respect of rights and freedoms by state administrations, territorial communities, public legal entities, as well as by all bodies carrying out a public service mission or by those that the Institutional Act decides fall within his remit*".[483] Lastly, the constitutional reform of 2008 instituted a direct form of constitutional recourse against an active legislation at the request of the *Conseil d'Etat* of the Court of Cassation, which makes the impugnation of a secondary legislation on constitutional grounds easier, as the previous mechanism relied mostly on the challenging by members of the parliament.[484]

---

[480] French 1946 Const. Preamble, France, art. 16
[481] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 102
[482] French 1958 Const, France, art. 1
[483] Ibid. art. 71
[484] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 103

For what concerns criminal law provisions on hate speech and discrimination, France can rely on a broad number of provisions that address discrimination and unprotected speech, from incitement to terrorism to incitement to hatred.[485] Regarding discrimination law, one of the most important provisions can be found within the labor code, in which article L 122-45 lists the grounds of prohibited discrimination.[486] The article lists the following grounds: origin, sex, manners, sexual orientation, age, family circumstances or pregnancy, genetic characteristics, belonging to an ethnic group or nation or race, political opinion, trade union activity, religious conviction, physical appearance, disability, or health condition.[487]

For what concerns racism, France has a unique approach to the issue, as the French doctrine refuses to recognize the concept of hate.[488] "*Origin and race are concepts that the law prohibits to take into consideration*".[489] However, this approach has been subject to extensive criticism, as the refusal to acknowledge this concept limits, for example, the scope of positive action on the matters of racial discrimination[490] and, more in general, undermines the effectiveness of legal protection against racial discrimination.[491] Nonetheless, many provisions in both the criminal code and the civil legislation deal with racism, racial discrimination, and incitement to discrimination.

Although many provisions of the French Penal and Civil codes can be applied to cases of hate speech, there is no official definition of the term in the French legislation. A commonly accepted definition within the French framework considers hate speech as referring to:

> [A] type of discourse that seeks to intimidate, incite violence or prejudice against a person or group of people based on various characteristics (race, age, gender, religion etc.). The term applies to written as well as verbal incitements as well as some public behavior.[492]

In the French civil legislation, the Press Freedom Act of 1889 is one of the most relevant legislations concerning hate speech, and it guarantees freedom of expression while simultaneously prohibiting

---

[485] Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States,* p. 17

[486] Lokiec, Pascal. "Discrimination Law in France", *University Paris XIII*, Available at: https://www.jil.go.jp/english/events/documents/clls08_lokiec.pdf, p. 1

[487] Ibidem

[488] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 114

[489] Ibidem

[490] Latraverse, S. 2011 *Report on measures to combat discrimination – Directives 2000/43/EC and 2000/78/EC*, Country report France, European Network of Legal Experts in the Non-discrimination Field, p. 120

[491] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 114

[492] Council of Europe, ELSA, *International Legal Research Group on ONLINE HATE SPEECH – Final Report*, (February 2014), p. 229

the publication of defamatory or insulting materials, considered damaging to the public order. The act states that:

> [a]nyone who, through utterances, slogans or threats in public places or meetings, or by way of written or printed material, […] directly incites the perpetrator(s) to commit the said offence, where that incitement is acted upon, shall be punished as an accessory.[493]

and that:

> [a]nyone who, by one of the means listed in Section 23, incites discrimination, hatred or violence towards a person or group of people on account of their origin, or their membership or non-membership of a particular ethnic group, nation, race or religion, shall be subject to a one-year prison sentence and a 300 000-franc fine, or one of these penalties only.[494]

Furthermore, article 24 of the act, the most relevant provision on hate speech in French legislation,[495] prohibits public incitement to hatred, discrimination, or violence on the basis of race, religion, nation, sexual orientation or gender identity, and article R- 624-3 of the Criminal Code broadens the prohibition of article 24, providing for the prohibition of private incitement as well.[496] Over the years the Freedom of the Press act has underwent some changes, such as the abolition of prison sentences for press offences or the expansion of penalties for defamation on the basis of race, religion or sexual orientation; but the general prohibition of offensive content remained present since the entry into force.[497] Thus, by employing this nonofficial definition along with the parameters that can be identified through international law provisions, national authorities and law interpreters are generally able to identify hate speech cases in the national framework.[498]

Now, the most used measure in the French penal code is a provision that concerns extremist speech and calls for the prohibition of "*direct provocation to terrorism or public apology of terrorism*".[499] The provisions on terrorism (provocation and apology) were introduced first in 1986 in the Law on

---

[493] Venice Commission, 2010. *Blasphemy, insult and hatred: finding answers in a democratic society*, Science and technique of democracy, No. 47, Council of Europe Publishing, p. 164

[494] Ibid. p. 165

[495] Council of Europe, ELSA, *International Legal Research Group on ONLINE HATE SPEECH – Final Report*, (February 2014), p. 228

[496] Ibidem

[497] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 107

[498] Ibid. p. 118

[499] Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States,* p. 17

the Freedom of the Press. They were later moved to the Penal Code's Chapter on terrorism in 2014, which now provides that the provocation and apology for terrorist acts is punishable by five years of imprisonment and a fine up to 75.000 euros.[500] It is interesting to note that in France the concept of apology of terrorism is defined in broad terms, and includes any type of expression that justifies, glorifies, expresses positive remarks towards terrorist acts. When asked to review the constitutionality of the provision, the French high court found the prohibition legitimate, necessary, and proportionate, as the act of apology to terrorism enhances the diffusion of potentially dangerous ideas that disturb the public order, which justifies the prohibition.[501] Moreover, the law of 2014 added an aggravating circumstance for the use of public means of communication to incite terrorism, justified by the "*particular magnitude of the broadcasting of prohibited messages that this mode of communication allows*".[502]

Lastly, for what concerns hate speech online, it can be considered punishable under the Freedom of the Press act, which prohibits incitement through the press or by any other mean of publication. However, this piece of legislation applies only to media professional and is rarely enforced by judges, who regard it as "*obsolete or excessive*".[503] Act 2004-575 of 21 June 2004 states that "*internet service providers and hosts have to contribute to the prevention of the dissemination of paedophiliac, revisionist and racist data*".[504] Furthermore, online content that provokes or incites terrorism can be blocked by authorities. Judges can order the withdrawal of content by service providers and, if the content is not removed, the blockage of the website.[505] To combat hate speech online the French National Assembly adopted a legislation on hate speech, referred to as the "Avia Bill", that imposed new obligations for online service providers. The legislation arose many concerns and was ultimately deemed unconstitutional by the French Constitutional Council. The details of the process will be the focus of the following paragraphs.

[500] Human Rights Watch website: https://www.hrw.org/news/2018/05/30/frances-creeping-terrorism-laws-restricting-free-speech [Accessed 20 January 2022]
[501] Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States,* p. 18
[502] Ibidem
[503] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 119
[504] Ibidem
[505] Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States,* p. 21

3.2.2 The "Avia Bill" on online Hate Speech: criticisms and concerns

The draft law *visant à lutter contre les contenus haineux sur internet*, commonly referred to as Avia Bill from the name of its principal proponent (Laetitia Avia, member of the National Assembly), was first submitted to the Parliament on March 20, 2019.[506] The proposition of law had the general purpose of combating hate speech online, and more specifically on the main social media platforms and search engines.[507] In fact, the proponents of the law expressed their concern for the expansion of the phenomenon of online hate speech, especially anti-Semitism and racism. The members of the parliament referred to a survey conducted in 2016 in which 70% of French citizens stated that they had already encountered instances of online hate speech, while 58% of citizens affirmed that they considered the Internet as the main instrument of hate speech diffusion.[508] Moreover, the proponents of the law reiterated that the phenomenon of hateful speech is "*too often tolerated*" when delivered in virtual form and that impunity is the rule for cyber-hate.[509] The Avia Bill aimed at translating the request of both the President Emmanuel Macron and the Prime Minister Édouard Philippe to "*place each stakeholder (platforms, users, internet providers) in front of their responsibilities within the committed fight against online hate*".[510]

For these reasons, the Avia Bill was proposed to the French parliament. The draft law aimed at combating the phenomenon of hate speech by imposing several obligations to internet providers. The first draft of the Bill proposed, among other things, that major online providers (identified by the number of users) would have been required to remove content inciting to hatred and insults on the grounds protected by French anti-discrimination law "*within 24 hours of notice*" by amending an already existing law, 2004-575 of 21 June 2004, and content inciting to terrorism and child pornography within 1 hour.[511] Moreover, the first iteration of the law required that the procedure of notification of illegal content would have been simplified by eliminating the requirement to describe the facts and reasons behind the reporting and included provisions on transparency obligations for companies and on the role of the French broadcasting regulator.[512]

Later, in May 2019, the Public Bill Committee adopted the draft law with a series of amendments which enlarged the scope of the bill by including a larger number of companies in its provisions,

---

[506] Proposition de loi 20 March 2019, n. 1785, *visant à* lutter contre *la* haine *sur* internet, Assemblée Nationale

[507] Library of Congress website: https://www.loc.gov/item/global-legal-monitor/2020-06-29/france-constitutional-court-strikes-down-key-provisions-of-bill-on-hate-speech/ [Accessed 22 January 2022]

[508] Proposition de loi 20 March 2019, n. 1785, *visant à* lutter contre *la* haine *sur* internet, Assemblée Nationale, p. 4

[509] Ibid. p. 4 - 5

[510] Ibid. p. 5

[511] Ibid. p. 6

[512] Article 19 website: https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/). [Accessed 23 January 2022]

removed the provision regarding the simplification of the notification procedure, criminalized malicious removal requests, specified the transparency requirements for companies, inserted an obligation to comply to the recommendations of the broadcasting regulator and criminalized companies failure to remove hateful content.[513] Article 19 recognized that this version of the draft law included many improvements for the protection of freedom of speech as the new version of the bill "has restored a degree of procedural fairness by, among other things, requiring individuals or companies who notify content to set out the facts and reasons for notifying such content" and includes "*provisions [...] aimed at improving transparency reporting on platforms or establishing internal complaints mechanisms*".[514] However, the organization expressed serious concerns about the draft law. According to their analysis, the draft law did not meet the international human rights law standards for the restriction of freedom of expression. The scope of the bill was considered too broad, as it would have bound a large number of companies and would have restricted numerous types of content. Moreover, the association expressed concerns in regard to the sanctions included in the bill, considered disproportionate under international law standards.[515]

Nonetheless, the final version of the draft law maintained the requirement to remove illegal content within 24 hours and the requirement to remove content containing child pornography and incitement to terrorism within 1 hour from the notification by the administrative authority to avoid breaching the law and risking 1 year of imprisonment and a fine of 250.000 euros.[516] The law was adopted by the National Assembly on the 9 July 2019 and awaited the examination by the senate, while the European Commission was notified of the draft law in August, as required by Directive (EU) 2015/1535 on laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (codification).[517]

The Avia bill was heavily criticized by many different subjects: members of the parliament, other countries, international organizations, and civil society organizations.[518] The Czech Republic sent a detailed opinion on the law, in compliance with the notification procedure of the EU,[519] followed by the request from the European Commission to postpone the adoption of the law, for addressing the

---

[513] Ibidem

[514] Article 19 website: https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/). [Accessed 23 January 2022]

[515] Ibidem

[516] Conseil constitutionnel [CC] [Constitutional Court] decision No. 2020-801 DC, June 18, 2020, Rec. France, para. 2

[517] European Parliament and Council Directive 2015/1535 of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (codification), L 241/1

[518] EDRi website: https://edri.org/our-work/frances-law-on-hate-speech-gets-thumbs-down/ [Accessed 24 January 2022]

[519] "*Where it emerges that the notified drafts may create barriers to the free movement of goods or to the free provision of Information Society services or to EU secondary legislation, the Commission and the other Member States may submit a **detailed opinion** to the Member State that has notified the draft*"
European Commission website: https://ec.europa.eu/growth/tools-databases/tris/en/about-the-20151535/the-notification-procedure-in-brief1/ [Accessed 11 February 2022]

possibility of a violation of the European E-Commerce Directive.[520] The main concerns and critical points of the law can be synthetized by analyzing the contribution by the European Digital Rights (EDRi) association, which includes 42 European human rights organizations and provided an analysis of the major critical points of the Avia Bill.[521] Regarding the single market, the association complained about the complexity that a multiplication of laws on online content entails. The French Avia Bill, similarly to the German's Network Enforcement Act, would further complicates the "*already confusing patchwork of national and European rules imposed on hosting providers that operate and offer services throughout the European Union*"[522] and would compromise the adoption of the Digital Service Act, mentioned in a previous chapter, which aims at creating a uniform framework of responsibilities and procedures for addressing hateful online content.[523]

The examination then focused on a more specific issue pertaining to the Avia Bill's contents. According to the association the bill would "*entrench the decision-making power of digital companies as to the legality of content with insufficient safeguards for the protection of users' rights to freedom of expression, due process and privacy, beyond France's territory*".[524] In fact, the timeframe for the removal of content is considered insufficient to allow companies to actually assess the severity and reality of the removal request. This, combined with the financial sanctions that would be imposed when failing to remove content, would lead major Internet Service Providers falling within the scope of the provision to delete all risky content in an attempt to avoid the fining, instead of really assessing the requests.[525] "*Overall, the foreseen provisions strongly encourage host providers to over-comply with notices and therefore, over-remove content, including legal content which is a risk for freedom of speech*".[526] Moreover, the limited time to assess the requests would encourage companies to use automated tools for the evaluation of content, which have been proven to be liable for failure as they do not take context into consideration.[527] In addressing these issues, the association referred to current European law that requires companies to remove content promptly, as well as the Charter of Fundamental Rights which requires limits on freedom of expression to be necessary and

---

[520] European Commission website: https://ec.europa.eu/growth/tools-databases/tris/en/index.cfm/search/?trisaction=search.detail&year=2019&num=412&mLang=EN [Accessed 23 January 2022]
[521] EDRi website: https://edri.org/about-us/ [Accessed 24 January 2022]
[522] EDRi, 18 November 2019, *Contribution to the examination of France's draft law aimed at combating hate content on the internet*, p. 7
[523] European Parliament website: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act [Accessed 18 January 2022]
[524] EDRi, 18 November 2019, *Contribution to the examination of France's draft law aimed at combating hate content on the internet*, p. 1
[525] Ibid. p. 2
[526] Ibidem
[527] Duarte, N., Llanso, E., Loup, A. 2017. *Mixed Messages? The limits of automated social media content analysis*, Center for Democracy and Technology, p. 5

proportionate.[528] "*We also remind legislators at the national and European level that Member States have a positive obligation to protect fundamental rights of individuals both offline and online. Delegating this legal obligation to private actors is wholly inappropriate*".[529] Similarly, French LGBT associations heavily criticized the bill for delegating the responsibility of contrasting hate speech online to private actors instead of working for facilitating access to judicial mechanisms that are already in place in the French framework and could be applied for online hate speech cases.[530]
Article 6 of the bill was also criticized. In this case, the provision stated that:

> When a judicial decision that has become final prohibits the total or partial reuse of content constituting one or more of the offences provided […] the administrative authority […] may request that the persons mentioned in Article 6(I), (1) of this Law, and any domain name provider, block access to any site, server or other electronic process giving access to content deemed unlawful by said decision.[531]

This measure effectively allows the French authorities to order the removal or blockage of websites that reproduce illicit content, referred to as mirror sites. A member of the EDRi Network, La Quadrature du Net, referred to this provision in a letter to the members of parliament which contained a juridical analysis of the Avia Bill, and asserted that giving such powers to police forces in the country would have entailed a high risk of political censorship, as it is almost impossible to stop the viral sharing of contents, companies would have no way of stopping the violation of the court judge by users, and the police would become the ultimate arbiter on which sites to obscure.[532]
The heavy criticism that characterized the drafting and adoption by the Parliament of the bill culminated in the referral of the Avia Bill to the Constitutional Council by a group of over 60 French senators, which put the adoption of the law by the Senate on hold.[533]

3.2.3 The decision of the French Constitutional Council

The French Constitutional Council is a French court established in 1958 through the Constitution of the Fifth Republic. Although it is not a hierarchically superior court compared to other French courts

[528] EDRi, 18 November 2019, *Contribution to the examination of France's draft law aimed at combating hate content on the internet*, p. 3
[529] Ibidem
[530] Liberation website: https://www.liberation.fr/debats/2020/01/21/feministes-lgbti-et-antiracistes-nous-ne-voulons-pas-de-la-loi-cyberhaine_1774297/ [Accessed 24 January 2022]
[531] Proposition de loi 22 Janvier 2020 n. 388, *visant à* lutter contre *la* haine *sur* internet, Assemblée Nationale, art. 6
[532] Le Quadrature du Net, 27 juin 2019, *Letter to the parliament - PPL Avia, refusez les mesures inutiles et dangereuses,* p. 3
[533] French Senate website: https://www.senat.fr/presse/cp20200619b.html [Accessed 24 January 2022]

(for example the *Cour de Cassation*), it is the judicial authority that has the power of reviewing and evaluating the constitutionality of French laws.[534] The Constitutional Council was called to evaluate the constitutionality of the provisions contained in the Avia Bill on online content by more than 60 senators, and the Council ruled on the matter on the 18 June 2020.

For article 1 paragraph 1, regarding the removal of child pornography and incitement to terrorism online within 1 hour, the senators argued that the provision would have been incompatible with the 2000/31 E-Commerce Directive of the European Parliament, that "*the interference with freedom of expression and communication would be disproportionate due to the lack of sufficient safeguards*".[535] Moreover, they affirmed that the required constraints imposed by the law on internet providers would have been impossible to satisfy and would consequently violate the principle of equality before the law.[536]

In addressing the concerns raised by the senators regarding the first paragraph of article 1, concerning the removal of content containing child pornography or incitement to terrorism within an hour, the Constitutional Council referred to article 11 of the Declaration of the Rights of Men and of the Citizen of 1789, which establishes the centrality and importance of the right to freedom of expression and applies to online services as well, as reiterated by the court.[537] The Council then referred to the French Constitution stating that, although the legislator can enact rules on the exercise of free speech, seen as the exercise of the right is "*a condition of democracy and one of the guarantees of respect for other rights and freedoms*",[538] any limitation to the right must be necessary, appropriate and proportionate. Lastly, the court acknowledged the legitimacy of the intent of the law, as the legislator had the intention of limiting and contrasting the diffusion of types of communication that are considered abusive of the right to freedom of expression (child pornography and apology of terrorism for instance). However, the requirements of the avia Bill would have entailed that the decision on what content to remove and deem unlawful would have been "subject to the sole discretion of the administration" and the limited time provided to the companies to remove content would have not allowed to appeal and obtain a decision by a judge before the enactment of sanctions.[539] For these reasons, the French Constitutional Council deemed the first paragraph of the Avia Bill unconstitutional.[540]

---

[534] French Constitutional Council website: https://www.conseil-constitutionnel.fr/en [Accessed 25 January 2022]
[535] Conseil constitutionnel [CC] [Constitutional Court] decision No. 2020-801 DC, June 18, 2020, Rec. France, para. 3
[536] Ibidem
[537] Ibid. para. 4
[538] Ibid. para. 5
[539] Ibid. para. 7
[540] Ibid. para. 9

For what concerns paragraph 2, regarding the obligation for companies to remove or obscure illegal content within 24 hours from the notification, the senators who requested the involvement of the court reiterated the concerns expressed regarding paragraph 1. The Court acknowledged that the provisions of the Avia Bill were aimed at ending the abuse of the exercise of freedom of expression by preventing the commission of acts that would have disturbed the public order, as well as avoiding the dissemination of such contents.[541] First, the court noted that the obligation to remove content within 24 hours in the context of the Avia Bill arose from the moment in which a user reports it. This entailed that there was no intervention by authorities or judges, but nonetheless the operator had to examine all the reported content to avoid being penalized.[542] Moreover, the Council emphasized that the examination of content would have been entirely responsibility of the company, even in cases which concerned legal technicalities or called for a contextual examination of the content. Seen as these types of examinations present some evident difficulties, the court affirmed that the 24 hours period provided for in the draft law for the removal of content seemed to be particularly short. Lastly, the court noted that the punishment, a fine of 250.000 euros, did not consider the possibility of repetitions in the reporting of contents nor did it provide exemptions from liability for online content providers in a reasonable manner.[543] For these reasons, considering the difficulties in assessing the illegal nature of the content in under 24 hours, the penalty, and the absence of causes for exoneration from liability, the Constitutional Council stated that the Bill would have encouraged the removal of content regardless of the illegality of it, thus impairing the exercise of the right to freedom of expression, and deemed paragraph 2 of article 1 unconstitutional.[544]

Having deemed article 1 unconstitutional, all the articles in the draft law that referred to the obligations established under said article were deemed unconstitutional as well, namely articles 4, 5, 7 and 18.[545] Thus, the central provisions of the Avia Bill were deemed unconstitutional, as they infringe on freedom of speech and are not necessary or proportionate to the aim, which put a definitive stop to the adoption of the law.

3.3 Italy and the limited scope of Hate Speech laws

The Italian framework of regulation on hate speech is, at the international level, similar to that of France. Italy has ratified all the major United Nations instruments mentioned in the previous

[541] Ibid. para. 13
[542] Ibid. para. 14
[543] Ibid. para. 17 - 18
[544] Ibid. para. 19
[545] Ibid. para. 26

chapters,[546] is a Member State of the EU, thus bound by its laws, and ratified the Council of Europe Convention on Cybercrime.[547] However, Italy did not ratify the additional protocol to the convention, which represents the first difference from the previously analyzed framework and is a first sign of the limitations of the Italian framework of regulation on hate speech.[548]

The Italian constitution, adopted in 1948, assigns to equality and fundamental freedoms a central role by including them in the first section of the text: the fundamental principles.[549] In fact, as stated by the Italian Ministry of Foreign Affairs in a communication to the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion of 2018:

> The Italian legal system aims at ensuring an effective framework of guarantees, to fully and extensively protect the fundamental rights of the individual. Indeed, we rely on a solid framework of rules, primarily of a constitutional nature, by which the respect for human rights is one of the main pillars.[550]

However, despite the central role that both equality and freedom of expression have in the Italian legal framework and the recent surge in hate speech cases in the country, Italy does not fully comply with international standards for what concerns hate speech regulation.[551] Italian criminal law prohibits the most serious forms of hate speech, but the exhaustive list of grounds does not include gender or sexual orientation. Moreover, the application of the already limited law on the matter is inconsistent.[552]

In Italy there is no specific law or regulation for hate crimes and hate speech,[553] thus this paragraph will center on the analysis of the Italian provisions that can be applied to hate speech cases, on the major national initiatives for combating the phenomenon and on the failed draft law on LGBTQ rights, DDL Zan.

---

546 Council of Europe Website: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=185 [Accessed 14 Jaunary 2022]

547 Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=185 [Accessed 18 January 2022]

548 Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=189 [Accessed 18 January 2022]

549 Italian 1947 Const, Italy

550 Ministry of Foreign Affairs and International Cooperation, Inter-ministerial Committee for Human Rights, *Italy's Remarks following communication from un Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on the red button protocol*, Mr. David Kaye (May 2018), p. 2

551 Article 19 website: https://www.article19.org/resources/italy-responding-hate-speech/ [Accessed 20 January 2022]

552 Ibidem

553 European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 147

3.3.1 Provisions directly and indirectly restricting hate speech

As mentioned previously, the Italian Constitution guarantees the protection of the right to freedom of expression in article 21, while the right to equality is referred to in articles 2 and 3 of the text. For freedom of expression and the right to equality in the media, the Gasparri law of 2004 includes the main principles that regulate the general structure of media broadcasting. More specifically, the law outlines the fundamental principles applicable to media services, namely the protection of freedom of expression of every individual, of freedom of opinion, of receiving or imparting information and ideas and the respect for the dignity of the person.[554] According to the law, media services must guarantee "*access to a large variety of information and contents, according to non-discrimination criteria*", "*the airing of programmes that respect the fundamental rights of the person*" and the banning of programs that "*contain incitement to hatred on any grounds or that, even with regard to the time of transmission, may harm the physical, psychological or moral development of minors*".[555] For what concerns equality, the main anti-discrimination law in the Italian framework is the 2003 Legislative Decree No. 215, through which the Italian Parliament ratified and included in the national legislation EU Directive 2000/43 on equal treatment irrespective of racial and ethnic origins.[556] The Legislative decree "*provides a definition of equal treatmet and expands the definition of discrimination*".[557] Moreover, the Decree includes an explicit definition of harassment, defined as "*[T]hose unwanted behaviours adopted on the grounds of race or ethnic origin that aim at or have the effect of causing the violation of a person's dignity and creating a hostile, intimidating, degrading, humiliating and offensive environment*".[558] Moreover, the decree creates the National Office Against Racial Discrimination (UNAR), one of the equality bodies that will be analyzed in the following paragraph. The legislative decree only directly refers to race and ethnic origins as protected grounds for discrimination; however, later ministerial guidelines have extended the mandate of the UNAR to deal with other types of discrimination as well, such as religion, personal opinions, disabilities, and sexual orientation.[559] Lastly, there are a few other relevant pieces of legislation that deal with equality and non-discrimination in employment and occupation and for what concerns judicial protection, namely Legislative Decree No. 216 of 2003, Legislative decree No. 5 of 2010, Legislative Decree No. 198 of 2006 and Law No. 67 of the same year.[560]

---

[554] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 18
[555] Legge 3 May 2004, n. 112, in G. U. May. 5, 2004, n. 82, art. 3 - 4
[556] Decreto Legislativo 9 July 2003, n. 215, in G. U. Aug. 12, 2003, n. 186
[557] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 19
[558] Decreto Legislativo 9 July 2003, n. 215, in G. U. Aug. 12, 2003, n. 186, art. 2
[559] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 20
[560] Ibidem

The Italian Criminal Code entered into force in 1931, during the fascist period; hence, there are no provisions criminalizing racism or other types of discrimination, except for article 415 which refers to social hate.[561] The criminalization of discrimination was introduced in the code through amendments, and the code now includes provisions that deal with the problem of discriminations and aggravating circumstances for crimes that include a discriminatory aspect.

Law No. 645 of 1952 prohibits propaganda for fascism and the reorganization of the fascist party, introducing the crime of apology of fascism in the Italian framework by stating that "*[a]nyone who promotes organizes or directs associations, movements or groups indicated in article 1, shall be punished with imprisonment from five to twelve years and a fine ranging from two to twenty million dollars*".[562] A subsequent law, No. 645 of 1975, introduced the crimes of racism and discrimination on the penal code by punishing with imprisonment the spread ideas of racial superiority or hatred and the incitement to discrimination or violence towards racial groups, "*unless the facts committed constitute a more serious crime*".[563] Therefore, the first instrument implementing the provisions of the CERD in the Italian framework did not include aggravating elements for discrimination.[564] However, the provisions of the law were modified by subsequent laws: the most important Italian legal instrument for the prosecution of racist and hateful acts, the Mancino law, and law No. 85 of 2006 regarding crimes of opinion.[565]

Law No. 205 of 1993, commonly referred to as the Mancino law, modified article 1 of the preceding instrument, and calls for penalties for incitement to the commission of violent acts or for the provocation on racial, ethnic, national, or religious grounds.[566] Article 1 was therefore modified, and now reads:

> Anyone who, by any means whatsoever, disseminates ideas based on racial or ethnic superiority or hatred, or commits or incites others to commit discriminatory acts on racial, ethnic, national or religious grounds, shall be subject to a maximum prison sentence of three years; […] anyone who, by any means whatsoever, commits or incites others to commit acts of violence or acts designed to provoke violence on racist, ethnic, national or religious grounds shall be subject to a prison sentence of six months to four years.[567]

---

[561] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 150

[562] Legge 20 June 1952, n. 645, in G. U. June, 23, 1952, n. 143, art. 2

[563] Legge 22 May 1975, n. 152, in G. U. May, 24, 1975, n. 136

[564] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 151

[565] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 22

[566] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 152

[567] Legge 25 June 1993, n. 205, in G. U. June, 26, 1993, n. 148, art. 1

Moreover, the law provides for a general aggravating circumstance for offences that are committed with the goal of spreading or inciting discrimination on racial, ethnic or religious grounds and states that:

> Where offences carrying a sentence other than life imprisonment are committed for reasons of ethnic, national, racial or religious discrimination or hatred, or for the purpose of facilitating the activities of an organisation, associations, movement or group pursuing these goals, the sentence shall be increased by half.[568]

It is important to note that the aggravating circumstance has a very large scope of application, as it can be applied to any type of offence (except those punishable with a life sentence), and included the new ground of religious discrimination.[569]
Law 85 of 2006 further amended the original law of implementation of the CERD in the Italian legislation by changing the description of the punishable behaviors.

> The law no longer punishes those who spread hate by any means, but, instead, those who promote ideas based on superiority or racial/ ethnic hatred; no longer those who incite, but those who instigate to commit or actually commit discriminatory acts based on racist, national, ethnic or religious grounds; no longer those who incite, but those who instigate to commit or actually commit violence or provocative acts to violence based on racist, ethnic, national or religious grounds.[570]

As a result, the primary legislation that concerns hate speech in Italy arises many concerns and critics. The protected characteristics are evidently limited, as the law only criminalizes hate speech on racial, ethnic, national, and religious grounds, leaving out gender, sexual orientation, disability, age and other grounds.[571] Moreover, the 2006 amendment limited the scope of the previous piece of legislation, by replacing the words "*spread by any means*" with the words "*promote*". In relation to this change, the Italian Supreme Court (*Corte di Cassazione*) stated that the new wording of the norm requires that the dissemination must be aimed at influencing the behavior of the audience and create

[568] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 152
[569] Ibid. p. 153
[570] Ibidem
[571] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 22

a concrete danger for the group to be considered sanctionable.[572] In the same sentence, the Court held that the substitution of the term "*incite*" with the term "*instigates*", "*perceived by some jurists as a more restrictive definition than 'incitement' in identifying the crime of 'hate speech'*",[573] had no real implication in the context of the law, and that the two terms are to be considered equivalent.[574] Finally, the court affirmed that:

> for the aggravating circumstance to be applicable, the internal motivation of the action is not sufficient; the expression has to present itself […] as intentionally aimed […] at raising in others feelings of racial hatred or, in any case, give rise […] to a practical danger of discriminatory behaviors.[575]

To partially supply for the lack of regulation on hate speech cases, Italian courts have applied a number of provisions that were not initially intended for hate speech but that can indirectly restrict it.[576] The crime of defamation, provided for in article 51 of the Italian Criminal Code, is often considered aggravated by racial or ethnic biases by Italian courts, especially in cases that could be defined as online hate speech. For example, in a 2015 case of the Court of Appeal of Taranto, the court ruled that an online comment stating that the Minister of Integration, Cecile Kyenge, should "go back to the jungle where she came from" amounted to a crime of defamation aggravated by racial hatred as the comment "*far from being a manifestation of freedom of speech – suggests the original inferiority of the person based in the color of their skin*".[577] Of a similar nature is the jurisprudence relating to the crime of threats, which tends to apply aggravating circumstances in cases in which the threats are accompanied by racial or ethnic hate.[578] Lastly, it is interesting to note that in a 2013 case, the Supreme Court applied the provisions on the crime of criminal conspiracy to a case pertaining the participation and promotion of an online neo-Nazi group whose aims included incitement to discrimination and violence.[579] In the case, the Supreme Court asserted that propaganda is more likely to be effective if it relies on new technologies, and that the virtual internet communities can be

---

[572] Cass., sez. III., 23 giugno 205, n. 2742/2015, Giur. It. 2015, para. 7
[573] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 22
[574] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 153
[575] Cass., sez. III., 23 giugno 205, n. 2742/2015, Giur. It. 2015, para. 10
[576] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 26
[577] Associazione per gli Studi Giuridici dull'Immigrazione website: http://www.asgi.it/banca-dati/corte-dappello-trento-sezione-penale-sentenza-del-1-giugno-2016/
[578] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 28
[579] Ibid. p. 26

considered organized criminal groups under the criminal code, which prohibits the promotion of organizations that have the aim of inciting to racial hatred.[580]

3.3.2 The role of equality institutions in Italy

To complete the overview on the Italian framework for hate speech cases, it is necessary to analyze some national initiatives and the role of equality institutions. In fact, under international law, it is required that states apply a range of positive measures for the promotion of the right to equality. Thus, equality institutions have proven to be an important and effective tool in implementing such measures in states.[581] For countering hate speech, equality institutions around Europe have typically undertaken the role of aiding and assisting victims, receiving complaints, and carrying out investigations.[582]

In Italy, two equality institutions can be identified in the context of combating hate speech. First, the National Office Against Racial Discrimination (UNAR)[583]. UNAR is the Italian institution tasked with ensuring that the right to equal treatment of all people, irrespectively of their race, ethnic origin, religion, age, sexual orientation, gender identity or disability is respected, in compliance with EU Directive 2000/43.[584] It is the main monitoring body in the Italian framework, providing two annual relations to the Italian government regarding discrimination in the country. Moreover, the institution gathers reports, carries out inquiries and awareness-raising campaigns, and formulates recommendations regarding specific phenomena or cases.[585] To facilitate these activities and the access for the public, UNAR provides a free and multi-language service, the Contact Center, that receives reports and testimonies regarding discriminatory facts and provides information and support for the victims of discriminatory acts.[586] The Contact Center staff includes two research experts on new media and social media, who are responsible for the Media and Internet Observatory of the UNAR, tasked with collecting data and analyzing the new forms of discrimination that are emerging in online media.[587] UNAR promotes the active development of an anti-racist culture through the creation of networks between NGOs, local and national government institutions, and civil society

---

[580] Cass., sez. III, 24 aprile 2013, n. 1126/2013, Giur. It. 2013, para. 4 – 5 – 6

[581] Article 19, 2018, *Responding to "hate speech": comparative overview of six EU countries*, Media Against Hate Campaign, p. 26

[582] Ibidem

[583] Ufficio Nazionale Antidiscriminazioni Razziali

[584] Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/che-cos-e-unar [Accessed 23 January 2022]

[585] Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/la-mission-di-unar [Accessed 23 January 2022]

[586] UNAR, Ufficio per la promozione della parità di trattamento e la rimozione delle discrminazioni fondate sulla razza o sull'origine etnica, *Relazione al parlamento sull'attività svolta e sull'effettiva applicazione del principio di parità di trattamento e sull'efficacia dei meccanismi di tutela*, 2020, p. 61

[587] Article 19, 2018, *Responding to "hate speech": comparative overview of six EU countries*, Media Against Hate Campaign, p. 27

associations, with the aim of "*building and strengthening a culture that contrasts any form of racism*".[588] The network created by UNAR relies on an IT system that allows the institution to access all the information and data gathered by local partners, which ensures a more effective monitoring by the institution.[589] Although UNAR plays a central role in combating discriminations by developing positive action policies, the institution is not able to bring cases of action or issue sanctions for cases brought to them or found by them. UNAR can, in some cases, promote informal reconciliation procedures between the parties.[590] In fact, when a report is considered relevant, i.e., when the office deems the description of the fact provided by the person who reported to be grounded and qualified as a racial discrimination, the office can start gathering relevant information pertaining to the case. The office can then facilitate informal conciliation between the two parties, provide a formal communication to the discriminating subject aimed at highlighting the issue and the possible solutions in compliance with the equality principle, and denounce the case to competent authorities.[591] It should be noted that in October 2020, most of the major ONGs and institutions dealing with hate crimes and hate speech in the Italian context, including UNAR, united in a national network, the *Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio*.[592] This network - which comprises several International NGOs, national institutions, researchers, and universities - aims at promoting monitoring mechanisms for hate speech and hate crimes, contrasting the phenomenon and the disinformation from which it stems, engaging in counter-speech and facilitating cooperation between subjects.[593]

The second equality institution in Italy is the Observatory for Security against Acts of Discrimination (OSCAD). The observatory was established in 2010 by the Chief of Police, Antonio Manganelli.[594] It unites representatives of all the police forces of Italy (*Polizia di Stato* and *Arma dei Carabinieri*) and plays an important monitoring role in the Italian framework, as the Observatory can receive reports and complaints regarding cases of discrimination and hate crimes through a dedicated hotline or through e-mails. The mission of this institution is to facilitate the reporting procedure of hate crimes, improve the monitoring effort and facilitate awareness raising within police forces.[595]

---

[588] Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/rete-antidiscriminazione [Accessed 23 January 2022]
[589] Ibidem
[590] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 36
[591] Ufficio Nazionale Antidiscriminazioni Razziali website: https://www.unar.it/portale/faq [Accessed 23 January 2022]
[592] Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/chi-siamo/ [Accessed 23 January 2022]
[593] Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/protocollo-d-intesa/ [Accessed 23 January 2022]
[594] Chirco, S. Gori, L. Esposito, I. 2020, "Quando l'odio diventa reato – Caratteristiche e normative di contrasto agli hate crimes", *Poliziamoderna*, p. 18
[595] Ibidem

Moreover, in 2013, the department of public security signed a Memorandum of Understanding through which OSCAD joined the TAHCLE program of the Office for Democratic Institutions and Human Rights (ODIHR).[596] The Training against Hate Crimes for Law Enforcement program is designed "*to improve police skills in recognizing, understanding and investigating hate crimes*" and allowed Italian police officers to be trained on the topic of hate crimes, enhancing their understanding of the phenomenon and of the best practices for dealing with discriminatory crimes.[597]
The two equality institutions, seen as they very similar mission statements, often cooperate in the Italian territory. For example, OSCAD and UNAR organized a series of workshops focusing on the prevention of hate crimes through which almost three-thousand National Police Officers received training.[598] In addition, according to a Memorandum of Understanding signed by the Department of Equal Opportunities and the Ministry of the Interior, UNAR has to report to OSCAD all "*'hate crimes' and racist acts of criminal relevance, including alleged 'hate speech' incidents reported to them*" while OSCAD is "*duty bound to report all discrimination cases that do not constitute crimes to UNAR for information and monitoring purposes*".[599] Thus, through these mechanisms, hate speech cases reported to OSCAD by UNAR are brought to the attention of the Postal Police section of the Police forces, which is the branch of the police that deals with all crimes committed online.[600]
It is important to note that in the monitoring report on the year 2020 provided by UNAR to the Italian Government, the institution referred to a surge in hate speech cases in the country identified through the analysis of the data gathered by the Contact Center, especially in online forums and social media.[601] In 2020, the institution analyzed 1002 cases, of which 913 were considered relevant. Of the relevant cases identified by the institution, 59.7% were related to discrimination on ethnic or racial basis, 20% to religious basis, 10.3% to sexual identity and gender, 5.4% to disability and architectural barriers, 2.7% to age, and 1.9% to various grounds.[602] In the report the office also provided a summary of four cases relating to specific issues, in an effort to highlight some of the key issues identified during the year. One of the cases relates to discrimination on the basis of sexual orientation, in which the CEO of a marketing company reported being verbally assaulted during a meeting with professional boxing federation's representatives at the CONI Palace of sport federations. The man

---

[596] Carabinieri website: http://www.carabinieri.it/arma/partners/osservatorio-per-la-sicurezza-contro-gli-atti-discriminatori-oscad [Accessed 24 January 2022]
[597] European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015, p. 163
[598] Ibidem
[599] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 36
[600] Ibidem
[601] UNAR, Ufficio per la promozione della parità di trattamento e la rimozione delle discrminazioni fondate sulla razza o sull'origine etnica, *Relazione al parlamento sull'attività svolta e sull'effettiva applicazione del principio di parità di trattamento e sull'efficacia dei meccanismi di tutela*, 2020, p. 62
[602] Ibid. p. 69

reported being addressed with specific insults relating to his sexual identity by one of the prominent members of the Italian Boxing Federation (the acting Vice President). The man contacted UNAR after filing a formal complaint through a legal representative, as the process seemed to be stalling. UNAR then proceeded to gather information and contacting the Boxing Federation, and in December 2020 the Secretary General of the federation sanctioned the author of the discrimination by suspending him from any activity related to the federation for a period of 80 days.[603] This analysis provides a first look into one of the critical points identified in the report: the lack of a framework of legislation for the protection of LGBTQ people in Italy. In fact, in a later section of the report, the UNAR refers to the "*necessity to fill this dangerous legislative void*" and provides an analysis of the draft law on homophobia, transphobia and discrimination on the basis of disability commonly referred to as DDL Zan, which will be analyzed in the following paragraph.[604]

3.3.3 Protection of LGBTQ people: the rise and fall of DDL Zan

The draft law on discrimination on the basis of gender, sexual orientation, and disability was proposed and approved at the first reading by the Italian parliament on 4 November 2020. The draft law was the result of a compromise between different proposals advanced in the previous year by members of the parliament (C. 107 Boldrini, C. 569 Zan, C. 686 Scalfarotto, C. 2171 Perantoni, C. 2255 Bertolozzi).[605] These proposals were developed to fill the void that can be identified in the Italian legal framework on hate speech and hate crimes. In fact, as underlined in the previously mentioned UNAR report, the Italian legislation is characterized by an evident lack of guarantees for what concerns the protection from discrimination and hate crimes based on sex, disability, sexual and gender identity, as the existing norms only refer to racial or ethnic discrimination.[606] Furthermore, the data gathered by the European Union Agency for Fundamental Rights (FRA) in 2020 highlighted the prevalence of everyday discrimination in Italy for LGBTQ people. According to the LGBTI Survey Data Explorer, 40% of the people interviewed reported being discriminated against based their LGBTQ identity, and almost 60% of Trans people interviewed reported feeling personally discriminated against based on gender identity.[607] Thus, the final version of the draft law DDL Zan was proposed to partially close the gap left by the current legislation.

[603] Ibid. p. 81
[604] Ibid. p. 172
[605] Ibidem
[606] Ibid. p. 173
[607] European Union Agency for Fundamental Rights website: https://fra.europa.eu/en/data-and-maps/2020/lgbti-survey-data-explorer [Accessed 25 January 2022]

The DDL was composed by ten articles which mostly aimed at modifying the articles 604-bis and 604-ter of the Penal Code, the previously mentioned Mancino law, to include discriminations "*based on sex, gender, sexual orientation, gender identity or disability*".[608] The first six articles of the draft law concerned the actual modification of the Penal Code, with article 1 defining the terms "sex", "gender", "sexual orientation" and "gender identity", and the following articles introducing the mentioned grounds in the articles of the penal code concerning hate crimes and aggravating circumstances for crimes. It should be noted that, to ensure the respect of the right to freedom of opinion, article 4 of the proposed law stated:

> For the purposes of this law, the free expression of beliefs or opinions as well as legitimate conduct attributable to the pluralism of ideas or the freedom of choices are excluded, provided that they are not suitable for determining the concrete danger of carrying out discriminatory or violent acts.[609]

The last three articles of the DDL were aimed at the implementation of positive measures for combating hate crimes and hate speech towards the LGBTQ community. Article 7 concerned the establishment of a national day for homophobia, lesbiphobia, biphobia and transphobia, article 8 instructed UNAR to develop a national strategy for the prevention and effective fighting of discrimination in the country, article 9 required the establishment of a 4 million euros found for active policies regarding human rights and equality, while article 10 required ISTAT, the Italian national statistical institute, to conduct a survey on discrimination and hate crimes every three years.[610]
Although apparently necessary, the draft law sparked some heavy criticisms both in the Italian parliament and in the public debate. The law, approved at the first reading by the parliament, encountered a series of difficulties and setbacks in its road to the second Italian parliament chamber, and was ultimately brought down by a secret ballot vote in the Senate.[611] One of the major critics moved to the DDL by those opposing it, other than the limitation to freedom of speech argument that most hate speech laws face, was that the modification of the law would have been superfluous, as the current legislation already prohibits hate crimes and could be, in theory, applied to cases of

[608] Senato della Repubblica, *Disegno di Legge (Misure di prevenzione e contrasto della discriminazione e della violenza per motivi fondati sul sesso, sul genere, sull'orientamento sessuale, sull'identità di genere e sulla disabilità)*, n. 2005, XVIII legislatura, 5 November 2020, art. 2
[609] Ibid. art. 4
[610] Wired website: https://www.wired.it/attualita/politica/2021/05/04/ddl-zan-articolo-spiegato/ [Accessed 25 January 2022]
[611] Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/10/28/ddl-zan-uno-stop-che-brucia-ma-che-non-ci-fermera/ [Accessed 25 January 2022]

homophobia and transphobia as well.[612] However, this critic has been countered through the analysis of the principle of equality, central in both International human rights law and in the Italian framework. Seen as the concrete application of the principle of equality requires a differentiation on the treatment of differing situation, it becomes necessary to "*provide an enhanced protection for individuals that are in conditions of greater weakness and vulnerability*".[613] Moreover, the argument is in clear contrast with regional human rights instruments, which require the general prohibition of discrimination,[614] and with EU provisions, which affirms that "*in defining and implementing its policies and activities, the Union shall aim to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation*".[615] Another critic moved to the DDL was the connection with the so-called "Gender theory" and its introduction in schools through the establishment of the national day against homophobia.[616] However, it is important to underline that the text of the draft law does not refer in any way to the mentioned theory, nor does it, in its final form, mandate the implementation of any type of activities within schools.[617]

Other than the explicit critics, the draft law also encountered some indirect oppositions by some members of the parliament. In fact, the calendarization of the reading at the Senate was postponed for almost six months by the Senate Justice Commission tasked with the first analysis of the text, which voted for the calendarization after a large mobilization of social media, of LGBTQ activists and NGOs.[618] However, after the calendarization, the draft law was met with harsh criticism by some of the senators, and, despite the numerous requests by political representatives and civil society subjects to have an open ballot,[619] was ultimately brought down by a secret ballot on the 27 October 2021.

Lastly, it should be highlighted that the path of the draft law was made even more challenging by the involvement of the Vatican in the process of approval. In fact, on the 17 June 2021, the Secretary for

---

[612] UNAR, Ufficio per la promozione della parità di trattamento e la rimozione delle discrminazioni fondate sulla razza o sull'origine etnica, *Relazione al parlamento sull'attività svolta e sull'effettiva applicazione del principio di parità di trattamento e sull'efficacia dei meccanismi di tutela*, 2020, p. 173

[613] Ibidem

[614] Council of the European Union, *Charter of Fundamental Rights of the European Union (2007/C 303/01)*, 14 December 2007, C 303/1, Available at: https://www.europarl.europa.eu/charter/pdf/text_en.pdf, art. 21

[615] Eur-lex website: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=IT [Accessed 25 January 2022], art. 10

[616] It is understood that this theory name refers to a theory that "*recognize gender as inextricably linked with social construct*" and "*challenge[s] the assumption that gender identity necessarily correlates with biological sex*" (United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/Issues/SexualOrientationGender/Pages/GenderTheory.aspx [Accessed 26 January 2022])

[617] Diritto.it website: https://www.diritto.it/ddl-zan-analisi-di-una-proposta-di-legge-molto-discussa/ [Accessed 26 January 2022]

[618] Wired website: https://www.wired.it/attualita/politica/2021/04/28/ddl-zan-calendarizzato-senato/ [Accessed 26 January 2022]
Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/03/29/legge-contro-l-omo-bi-transfobia-la-misoginia-e-l-abilismo-a-quando-la-calendarizzazione-in-senato/ [Accessed 26 January 2022]

[619] Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/10/26/ddl-zan-voto-senato-sia-palese/ [Accessed 26 January 2022]

Relations with States of the Vatican employed diplomatic means to explicitly take a stand against some of the provisions in the draft law. In fact, the Secretary delivered a "verbal note" to the Italian ambassy at the Vatican asking the Italian government to modify the DDL, as some of the provisions allegedly violated the agreement between the Italian government and the Vatican state of 1929.[620]

3.4 United States' liberal vision on Hate Speech: The First Amendment

As mentioned previously, the United States of America have a completely different approach to the issue of hate speech compared to other western democracies, especially in comparison with EU laws and standards. In fact, *"[a]gainst [the] global and European trend, the United States stands as a fierce dissenter*".[621] In the US, the First Amendment to the Constitution is the central instrument protecting and ensuring freedom of speech, religion and of the press.[622] Starting from the 1960s the US Supreme Court has developed a consistent jurisprudence on free speech based on the First Amendment, generally allowing "*more rather than less dissenting speech*",[623] and concluding that regulation on speech, even in cases in which the speech could damage individual dignity, would be unconstitutional.[624]

This paragraph will analyze the history and interpretation of the US First Amendment and the jurisprudence of the US Supreme Court on free speech developed in the last fifty years. Lastly, the analysis will focus on a landmark decision on freedom of speech and hate speech, the 2011 case of Snyder v. Phelps.

3.4.1 The First Amendment: history and first interpretations

The First Amendment to the Constitution of the US is inserted in the context of the Bill of Rights, adopted in 1791. Originally aimed at limiting the power of the Federal Government of the United States, the provisions of the mentioned text now apply to all US governments, federal and local

---

[620] Corriere della Sera website: https://www.corriere.it/cronache/21_giugno_22/vaticano-ddl-zan-legge-testo-b13294ba-d2d0-11eb-9207-8df97caf9553.shtml [Accessed 26 January 2022]

[621] Heinze, Eric. "Viewpoint Absolutism and Hate Speech." *The Modern Law Review*, vol. 69, no. 4, 2006, pp. 543–582, p. 544

[622] History website: https://www.history.com/topics/united-states-constitution/first-amendment#:~:text=The%20First%20Amendment%20to%20the,speech%2C%20religion%20and%20the%20press.&text=The%20amendment%20was%20adopted%20in,civil%20liberties%20under%20U.S.%20law [Accessed 28 January 2022]

[623] Boch, Anna. 2020, "Increasing American Political Tolerance: a framework excluding hate speech", *Socius: Sociological Research for a Dynamic World*, Vol. 6, p. 1

[624] Stone, Geoffrey. 1994, "Hate Speech and the U.S. Constitution", east European Constitutional Review, Vol. 78, p. 78

alike.[625] The first iteration of the speech and press provision of the Bill of Rights, proposed by the original drafter of the bill James Madison, stated "*[t]he people shall not be deprived or abridged of their right to speak, to write, or to publish their sentiments; and the freedom of the press, as one of the great bulwarks of liberty, shall be inviolable*".[626] However, after the reading at the House of Representatives, the language of the provision was modified first by the special committee and later by the Senate, which changed the provision to read:

> That Congress shall make no law abridging the freedom of speech, or of the press, or the right of the people peaceably to assemble and consult for their common good, and to petition the government for a redress of grievances.[627]

The religion clause was later added by the Senate, and thus the final wording of the provision, which remained the same since its adoption, states:

> 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.[628]

It should be noted that the simple wording of the First Amendment was an intentional choice of the legislators. In fact, James Madison stated, referring to the provisions of the Bill, that the main dangers for the Bill would arise "*from discussing and proposing abstract propositions*" and that "*if we confine ourselves to an enumeration of simple acknowledged principles, the ratification will meet with but little difficulty. Amendments of a doubtful nature will have a tendency to prejudice the whole system*".[629] Therefore, the provisions of the Bill of Rights are deliberately simple and brief, which facilitated the adoption at the time, but enhanced the difficulties in applying and interpreting them afterwards.[630] Initially, it is possible that the First Amendment was interpreted in a way that could be compared to that of modern Western European states. In fact, in the context of the approval of the provision, it is likely the view of the legislators was aligned to the common law view of the time on

---

[625] Cogan, N. H. 2015, *The Complete Bill of Rights: The Drafts, Debates, Sources, and Origins*, II ed., New York, Oxford University Press, preface
[626] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]
[627] Ibidem
[628] U.S. 1789 Const. Amend. 1, U.S.
[629] Cogan, N. H. 2015, *The Complete Bill of Rights: The Drafts, Debates, Sources, and Origins*, II ed., New York, Oxford University Press, p. 274
[630] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]

the matter.[631] Hence, it can be said that the initial intention of the provision, and the first interpretations, considered the right to freedom of the press as consisting in laying "*no previous restraints upon publications*",[632] but when the publication "*shall on a fair and impartial trial be adjudged of a pernicious tendency, [the punishment] is necessary for the preservation of peace and good order, of government and religion, the only solid foundations of civil liberty*".[633] Consequently, by following this common law principle, "*the will of individuals is still left free; the abuse only of that free will is the object of legal punishment*".[634]

However, despite the probability of this hypothesis, it appears that in the aftermath of the Sedition Act the interpretation of the First Amendment started to move towards a more absolute connotation of the right to freedom of speech, as interpreting the provision became more difficult.[635] The act, passed by the House of Representatives in 1798, permitted the deportation, fining or imprisonment of anyone who published "*false, scandalous, or malicious writing against the government of the United States*".[636] The unpopularity of the provision and of the administration that passed it resulted in the loss of the 1800 election, which favored Thomas Jefferson (a fierce opposer of the act), and the Sedition Act expired in 1801.[637] Although the Act was short lived, the debate that ensued because of its approval helped shaping the following arguments for the constitutional protection of freedom of speech in the US.[638] In fact, seen as the debate over the Sedition Act took place so shortly after the adoption of the Bill of Rights, when those who drafted the law were still the representatives, it becomes difficult in modern times to effectively discern the original intentions of the legislators. An interesting interpretation of the provision can be drawn from the analysis of what a "right" was in the rationale of the years in which the provision was drafted. At the time of the adoption, American legislators understood rights as being divided into natural rights and positive rights.[639] Natural rights were intended as "*all the things that we could do simply as humans, without the intervention of a government*",[640] but were not yet characterized by the modern definition of being a legally enforceable

[631] Ibidem
[632] Blackstone, W. 1979, *Commentaries on the Laws of England: A Facsimile of the First Edition of 1765--1769*. Chicago, University of Chicago Press.
[633] Ibidem
[634] Ibidem
[635] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]
[636] History, Art and Archives United States House of Representatives website: https://history.house.gov/Historical-Highlights/1700s/The-Sedition-Act-of-1798/ [Accessed 28 January 2022]
[637] Ibidem
[638] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]
[639] Campbell, Jude. 2018, "What did the First Amendment originally mean?", *Richmond Law Magazine*, available at: https://lawmagazine.richmond.edu/features/article/-/15500/what-did-the-first-amendment-originally-mean.html

[640] Ibidem

privilege. Thus, natural rights "*could be restricted by law to promote the good of the society*".[641] From this, it is possible to assume that restrictions to the right to free speech were considered possible if they were necessary to promote the public wellbeing. However, as demonstrated by the disagreement on the Sedition Act, legislators did not agree on what limitations were necessary, even when understanding natural rights in the same manner.[642]

Even though it is difficult to identify the drafters' intentions concerning freedom of speech, it is undeniable that, starting from the early twentieth century, the jurisprudence, and the general understanding that the provision on freedom of speech applies to both prior and subsequent punishment started to grow, becoming more and more prominent after the 1960s.[643]

3.4.2 US Supreme Court Jurisprudence

To understand the US framework on the regulation of hate speech cases, analyzing the jurisprudence of the Supreme Court is essential, as the judicial interpretations of the First Amendment are the primary source of law in these cases.[644] It should be highlighted that in the US the pervasiveness of this particular constitutional right, freedom of speech, goes beyond the mere protection awarded by the Constitution. "*Americans have a deep-seated belief in free speech as a virtually unlimited good and a strong fear that an active government in the area of speech will much more likely result in harm than in good*".[645] This strong conception of the right to freedom of speech stems from a variety of factors, which include the preference for liberty over equality and a strong commitment to individualism, and it could be said that "*in essence, free speech rights in the United States are conceived as belonging to the individual against the state*".[646] In general, the jurisprudence of the US Supreme Court of the last century reflects this conception of the right.

The first group of cases that addressed the issue of hate speech, more precisely of incitement to violence, have been referred to as "the World War I cases".[647] All decided in the same year, the four cases involved advocacy of resistance to military conscription and distribution of leaflets against the involvement of the US in the anti-Bolshevik fight in Russia. In this context, Judge Holmes first

[641] Ibidem
[642] Ibidem
[643] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]
[644] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 464
[645] Rosenfeld, Michel. 2003, "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis ", *Cardozo Law Review*, p. 1530
[646] Ibid. p. 1529
[647] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 471

referred to what would become the clear and present danger test (Schenck Test). In fact, in his opinion on the Schenck v. U.S. case regarding the distribution of leaflets in an attempt to obstruct the recruiting system, in violation of the Espionage Act of 1917.[648] Judge Holmes affirmed:

> The question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent.[649]

The primary holding of the case thus stated that "*if speech is intended to result in a crime, and there is a clear and present danger that it actually will result in a crime, the First Amendment does not protect the speaker from government action*" and the case met the set standard.[650] The same reasoning was applied in two other similar cases concerning violations of the Espionage Act, Frohwerk v. U.S. and Debs v. U. S., as the acts committed by the accused were considered likely to result in the commission of crimes, and the involvement of the US in the war made the absence of an actual obstruction to the government irrelevant.[651]

The fourth case, Abrams v. U.S., concerned the distribution of leaflets that denounced the United States' involvement in the aftermath of the Russian Revolution and advocated for "*a general strike by workers to force the U.S. to keep its armies out of Russia*".[652] In this instance the Court upheld the conviction through the clear and danger present test set out in the Schenck case.[653] However, Judge Holmes dissented on two grounds: first, the Judge argued that the purpose of the leaflets did not show any specific intent to hinder the war efforts, as the war was waged against Germany and the leaflets concerned the ant-Bolshevik intervention; second, Holmes affirmed that, even when assuming ill-intent by the accused, the actions did not present any actual danger, thus failing the clear and present danger test.[654] The dissenting opinion expressed by Judge Holmes helped understanding three central elements of the clear and present danger test for limiting freedom of expression, which still apply today:

[648] Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/249/47/#tab-opinion-1928047 [Accessed 29 January 2022]
[649] Schenck v. U.S., 249 U.S. 47 (1919)
[650] Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/249/47/#tab-opinion-1928047 [Accessed 29 January 2022]
[651] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 471
[652] Ibidem
[653] Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/250/616/ [Accessed 29 January 2022]
[654] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 472

> (1) that specific intent to bring about the forbidden result is a prerequisite for suppression of the speech by the criminal law; (2) that the intent must be to bring about the forbidden result immediately; and (3) that substantial punishment for such intent requires an objective likelihood of success under the circumstances.[655]

It should be noted that the requirement of imminence set out in this opinion was later found in a majority opinion on the case of Brandenburg v. Ohio, which reinforced the legitimacy of the test.[656] Therefore, it is possible to state that the American legal system does provide for some limitations to the right to free speech, provided they fall within the standards set by the above-mentioned principles. Another jurisprudential principle that limits, to a certain extent, freedom of expression is what Rubenfeld calls "the anti-orthodoxy principle".[657] The statement that best summarizes this principle was delivered by Judge Jackson in the case of West Virginia State Board of Education v. Barnette of 1943:

> If there is any fixed star in our constitutional constellation, it is that no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion, or other matters of opinion or force citizens to confess by word or act their faith therein. If there are any circumstances which permit an exception, they do not now occur to us.[658]

In other terms, the anti-orthodoxy principle states that a state cannot decide on what types of expressions can or cannot be expressed, although it is possible for the state to prescribe what to say in specific situations.[659] As explained by Rubenfeld

> while it is true that "there is no such thing as a false idea" under the First Amendment there is clearly such a thing as a false fact. The laws of libel, fraud, perjury, and so on, all punish people for speaking falsely on matters of fact. [660]

Thus, in the US it is possible to punish false statements when concerning matters of fact (a doctor can be punished for expressing an opinion if they express it in a negligent way).[661] However, this principle creates a set of difficulty, especially in identifying which expressions are to be considered factual,

---

[655] Ibidem
[656] Ibid. 475
[657] Rubenfeld, J., 2001. The First Amendment's Purpose. *Stanford Law Review*, vol. 53, p. 818
[658] Ibidem
[659] Ibid. p. 819
[660] Ibidem
[661] Ibid. p 818

and therefore in theory punishable, and which are mere statements of an opinion. The current doctrine "*solves this problem to some extent by holding that government can punish speech as false only when it can prove false*",[662] hence relying on a burden of proof. To summarize, the current understanding of the principle entails that "*there can be no blasphemy in American law, but there can be lies, misrepresentations, failures to disclose material information, breaches of confidentiality, and so on*".[663]

It is now necessary to understand the approach of the Supreme Court to cases of speech conduct, which is generally guided by the O'Brien test. The case of U.S. v. O'Brien of 1968 concerning the burning of a draft card to express a dissent is the origin of the four-part test that is generally applied to similar cases, as the case remains the leading one on the matter. In the case, the Court affirmed that to judge a law on nonverbal expression is legitimate:

> [1] if it is within the constitutional power of the Government; [2] if it furthers an important or substantial governmental interest; [3] if the governmental interest is unrelated to the suppression of free expression; and [4] if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest.[664]

In the O'Brien case, the court held that the federal government could punish the speech act, not because of the destruction, but because of the function of the destroyed object. Hence, the destruction of a draft card, which has a function in the draft system of the U.S., could be sanctioned,[665] "*since the government has an important interest in an effective draft system, the First Amendment does not void a law against burning draft cards, especially since the act of burning a draft card does not implicate a substantial speech interest*".[666] Contrarily, in the case of Texas v. Johnson of 1989 for example, concerning the burning of an American flag, the court held that the destruction of the flag could not be prohibited by states nor by the federal government, as the object was privately owned, and the burning did not jeopardize any important state interest.[667]

Lastly, it is necessary to analyze the Spence test on nonverbal conduct. This test, first applied in the case of Spence v. Washington of 1974, is based on the Court ruling stating that:

---

[662] Ibid. p. 820
[663] Ibid. p. 821
[664] Ibid. p. 770
[665] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 469
[666] Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/391/367/ [Accessed 29 January 2022]
[667] Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, p. 470

> [I]n deciding whether particular conduct possesses sufficient communicative elements to bring the First Amendment into play, we have asked whether 'an intent to convey a particularized message was present, and whether the likelihood was great that the message would be understood by those who viewed it.[668]

The Spence test is, de facto, a prerequisite for the application of the already mentioned O'Brien test. In fact, "*a person trying to invoke the O'Brien test has to satisfy Spence*".[669] However, it should be noted that Spence test is regarded as unsatisfactory by some scholars,[670] as it does not, for example, consider art expressive enough to trigger the protection of the First Amendment.[671]

To conclude the analysis, it can be said that the US framework for freedom of speech is one of the most limited when concerning possible limitations to speech and hate speech cases in the US reflect this position. In fact, in the case of Brandenburg v. Ohio of 1952, which set the standard for the current approach of the Court, the Court ruled in favor of a Ku Klux Klan leader who had made a speech containing anti-Semitic and anti-black statements, alluded to a possible revenge and had been convicted by the state of Ohio.[672] In this case, the Supreme Court decided against the conviction, setting aside the criminal sanction of the accused, as "*the Klan may have advocated violence, but that it had not incited it*",[673] and thus the conviction did not pass the Schenck test.

3.4.3 Snyder v. Phelps, a landmark decision on Freedom of Speech

The analysis of the case of Snyder v. Phelps (2011) provides an interesting addition to the abovementioned US framework on hate speech. In fact, the decision of the Court on this case created a new precedent for the intentional creation of tragic public spectacle by awarding the activities of the Westboro Baptist Church the First Amendment protection.[674]

To understand the case, a premise on the controversial nature of the church's activities is necessary. Founded in 1955 by Fred Waldon Phelps, the defendant in the case at question, the Westboro Baptist Church is situated in Topeka (Kansas) and is composed primarily by members of the Phelps family. The church became known to the public in 1998, when the church picketed the funeral of Matthew

---

[668] Rubenfeld, J., 2001. The First Amendment's Purpose. *Stanford Law Review*, vol. 53, p. 772
[669] Ibid. p. 772
[670] Ibid. p. 773
[671] Ibidem
[672] The Free Speech Center Website: https://www.mtsu.edu/first-amendment/article/189/brandenburg-v-ohio [Accessed 29 January 2022]
[673] Rosenfeld, Michel. 2003, "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis ", *Cardozo Law Review*, pp. 1537
[674] Lane Burner, M., Balter-Reitz, Susan. 2013, "Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle", *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 653

Shepard,[675] a gay college student murdered in Fort Collins (Colorado) carrying signs covered in homophobic slurs.[676] (in nota: Mattew Shepard was an American college student who was severely beaten because of his sexual orientation and was left to die in 1998. On October 7, 1998, Shepard was befriended by two men, Aaron McKinney and Russell Henderson, who were posing as gay in order to lure him away from a local bar. They drove him to a rural area where they tied him to a fence, administered a brutal beating, and left him to die in the cold. He was discovered and hospitalized, though he succumbed to his injuries. "*Because of their intense, religiously based hatred of homosexuality, U.S. military policy permitting—if not condoning—homosexual behavior, the Catholic Church, Jews, and other groups whose beliefs differ from their own*".[677] As explained by a former member of the church, the church "*had taken to the streets because [it] had a solemn duty to obey God and to plead with [its] neighbors to do the same. It didn't matter that the world hated the message*".[678] Hence, it is possible to state that the church explicitly searches for media attention, irrespectively of it being a positive or negative representation, and that their protests are "*strategically orchestrated for maximum media exposure*".[679]

The Snyder v. Phelps case concerned the picketing of a soldier's funeral, Matthew Snyder, by a group of protesters from the aforementioned church, guided by Fred Phelps. The picketing took place on public grounds, "*approximately 1,000 feet from the church where the funeral was held*",[680] and the protesters peacefully showed their signs, which contained vitriols of various kinds directed at soldiers, homosexuals, and America in general.[681] After the funeral, the father of the soldier filed a diversity action against the church, Fred Phelps and his daughters.

> Snyder's claims […] were grounded in an implicit theory of the private sphere, arguing emphatically that a human being's private suffering and grief should not be used in a mercenary fashion as a public platform for religious or political grand- standing by another party.[682]

[675] "Mattew Shepard was an American college student who was severely beaten because of his sexual orientation and was left to die in 1998. […] On October 7, 1998, Shepard was befriended by two men, Aaron McKinney and Russell Henderson, who were posing as gay in order to lure him away from a local bar. They drove him to a rural area where they tied him to a fence, administered a brutal beating, and left him to die in the cold. He was discovered and hospitalized, though he succumbed to his injuries." Britannica website: https://www.britannica.com/biography/Matthew-Shepard [Accessed 30 January 2022]

[676] BBC website: https://www.bbc.com/news/world-us-canada-45968606 [Accessed 30 January 2022]

[677] Lane Burner, M., Balter-Reitz, Susan. 2013, "Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle", *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 655

[678] Phelps-Roper, M. 2019, *Unfollow A memoir of loving and leaving the Westboro Baptist Church*, New York, Farrar, Straus and Giroux, p. 44

[679] Lane Burner, M., Balter-Reitz, Susan. 2013, "Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle", *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 657

[680] Snyder v. Phelps, App. No. 09 – 751 (SCOTUS, 2 March 2011), p. 1

[681] Ibidem

[682] Lane Burner, M., Balter-Reitz, Susan. 2013, "Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle", *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 660

The district court dismissed the defamation and publicity given to private life charges, arguing that the religious opinions expressed by the church could not be considered defamatory and that the information delivered by the church was already public.[683] However, the church was considered guilty on the other grounds (intentional infliction of emotional distress (IIED), intrusion upon seclusion, and civil conspiracy)[684] and held liable for compensation of 2.9 million dollars by a jury. The church challenged the verdict "*as grossly excessive and sought judgment as a matter of law on the ground that the First Amendment fully protected its speech*".[685]

The case was then referred to the Court of Appeal, which overturned the original judgement and afforded the church the protection of the First Amendment, and ultimately to the Supreme Court, who ruled in favor of the church.[686] The Supreme Court final decision held that the expression in question was entitled to First Amendment protection, as the church was speaking on matters of public concern, which could not be proven to be false and were only characterized by "hyperbolic rhetoric".[687] "*The Court affirmed the principle of content neutrality and held that there was a place for vitriol in public debate, even if it caused various hurts harms, and even if it made little contribution to that debate*".[688] Thus, even in a case in which the vitriolic hate speech was so evident, the Supreme Court afforded the expression the protection of the First Amendment.

The central issue that is highlighted in the analysis on the Snyder v. Phelps case concerns the public concern test applied in the case. The test, drawn upon in the decision, affirms that courts should considered the entire context and form of the speech at question, as public concern is "*a subject of general interest and of value and concern to the public at the time of publication*".[689] "*A statement's arguably inappropriate or controversial character . . . is irrelevant to the question whether it deals with a matter of public concern*".[690] What is contested by some scholars is the lack of consideration by the Supreme Court of the entire context of the speech. In fact, in applying the public concern test, it is undeniable that the speech uttered by the church can be seen as public speech, as public speech is any speech that attracts attention.[691] However, according to some analysis, the Court did not take into consideration the complete picture of the case, namely the "*impact of both television and Internet*

[683] Ibid. p. 661
[684] Ibid. p. 660
[685] Ibidem
[686] Delgado, R. Stefancic, J. 2018, *Must we defend Nazis? – Why the First Amendment Should not Protect Hate Speech and White Supremacy*, New York, New York University Press, p. 19
[687] Snyder v. Phelps, App. No. 09 – 751 (SCOTUS, 2 March 2011), p. 2
[688] Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, pp. 396
[689] Snyder v. Phelps, App. No. 09 – 751 (SCOTUS, 2 March 2011), p. 2
[690] Ibidem
[691] Lane Burner, M., Balter-Reitz, Susan. 2013, "Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle", *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 669

coverage on the Snyder family and their reputation" and chose to concentrate only on the primary act of the speech, the signs at the funeral, "*limit[ing] their analysis of the message to the content that was least personal and least impactful on the Snyders*".[692]

Nonetheless, the approach of the Court to the Snyder v. Phelps case is consistent with the jurisprudence on similar types of speech (Collins v. Smith) and falls within the general framework of the "argument from democracy", used to protect free speech on four grounds that connect to democracy, as public discussion ensures the legitimacy of the law, communication provides citizens with the information needed to make choices, the protection of speech allows citizens to criticize the government and communication is necessary for autonomy.[693]

> While many of the other free speech defenses have come under considerable critical scrutiny, there is widespread consensus in the free speech literature and jurisprudence that the argument for speech protection on the grounds of democracy or political communication is the most compelling.[694]

The Snyder v. Phelps judgment is a clear example of the application of these reasoning by the US Supreme Court to hate speech cases and speech in general.

---

[692] Ibid. p. 670

[693] Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, pp. 309

[694] Ibidem

Conclusion

This work aimed at analyzing the phenomenon of hate speech by focusing first on the main international standards and regulations allowing the limitation of freedom of speech, in an attempt to provide a complete framework of the legal instruments available and of the main challenges that arise from the concept of hate speech. In the first chapter, the focus on the main provisions and international law instruments allowed for the drawing of certain conclusions. Although there is no clear standard for hate speech in international law, the limitation of the right to freedom of speech is provided for in many different international and regional conventions. In fact, even though the final text does not include a specific provision limiting freedom of expression, the risks of unregulated speech were addressed by the drafters of the Universal Declaration of Human Rights,[695] and starting from the International Covenant on Civil and Political Rights the limitations to the right became explicit.[696] Thus, by considering the instruments limiting the right to free speech in conjunction with the anti-discrimination provisions, it can be said that international law does provide for a level of punishability and criminalization of certain forms of expression, even if the term hate speech is not used directly. Moreover, the United Nations have shown a specific interest in filling the void that created around what hate speech is and how it should be addressed. With the Rabat Plan of Action and the recommendations of the Special Rapporteur on Freedom of Opinion,[697] the institution has provided a needed clarification on the international standards for the prohibition of certain kinds of expression, giving specific indications to states on how to effectively implement the provisions of the main human rights conventions.[698] As a result, at a European level, two actors have actively promoted the unified application of international law standards on the limitations to free speech, the Council of Europe and the European Union, while the European Court of Human Rights concretely applied the provisions, underlying that the limitations are provided for by the European Convention on Human rights and, more in general, by international law.[699] At the Interamerican level, the concrete application of the provisions has proven difficult; nonetheless, the Interamerican Commission on Human Rights has highlighted that, although in principle all expression is protected by the right to free speech, the right

[695] Farrior, Stephanie. 1996. "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, pp. 1-98, p. 12-13
[696] UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, art. 19, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx
[697] Ibidem
[698] UN Human Rights Council, *Annual report of the United Nations High Commissioner for Human Rights : Addendum, 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*, 11 January 2013, A/HRC/22/17/Add.4, para. 17 – 18, available at: https://www.ohchr.org/en/issues/freedomopinion/articles19-20/pages/index.aspx
[699] European Court of Human Rights, *Factsheet - Hate speech*, Press Unit, June 2020

is not absolute.[700] Thus, it can be said that at the international level there is a level of consensus on the need to limit freedom of speech in some specific cases, but the lack of clear and agreed upon standards for what hate speech is and how it should be addressed remains, as there has been no real uniform advancement in the legislation.

This lack of standards and consensus has been further analyzed in chapter two, which highlighted the main challenges that regulating hate speech poses. Other than providing an analysis of the ideological debate, the chapter underlined how the phenomenon of hate speech is real and widespread, especially in online setting and towards specific communities. Social media have proven to be a central tool in the spreading of hate speech, and the difficulty in addressing the issue of online hate has emphasized the incompleteness of the current international framework for extreme speech cases. In fact, to properly combat the phenomenon, a multilateral agreement would be needed.[701] However, as of today such an agreement has been impossible to achieve, for the differences in the application of the international limiting norms on speech and in the ideological positions on the limitations have nullified the attempts. It can be said that the lack of clear standards combined with the diffusion of the Internet and social media has fueled the rise in cases of online hate. Moreover, the impossibility to hold the IT providers liable for what happens on online platforms and the inconsistencies in their community standard enforcement has resulted in a general inability to address the issue of online hate speech.[702] Lastly, by analyzing the prevalence of certain forms of hate speech, namely hate speech based on sex or gender, and the output of the main international court dealing with such cases, the European Court of Human Rights, it can be affirmed that there is a general lack of application of the norms that protect from such speech. In fact, the number of cases that deal with sexist or homophobic hate speech are not reflective of the incidence of the phenomenon.[703]

Finally, the third and last chapter has tried to provide an insight on four national systems of legislation on hate speech, highlighting the different approaches and the connection with the aforementioned issues. In the German case, the approval of the new law on social media responsibility for hate speech has highlighted some of the major critics that are moved to such kinds of laws, although the effectiveness of the provision has not been fully understood yet. In the French case, the final verdict of the Constitutional Council on the Avia Bill and the critics that the bill has received[704] showed that

[700] IACHR, *Annual Report of the office of the Special Rapporteur for Freedom of Expression*, Office of the Special Rapporteur for Freedom of Expression, OAS. Official records; OEA/Ser.L/V/II, (2015)

[701] Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 234

[702] Siegel, A. 2020, *Online Hate Speech,* in Persily, N. Tucker, J ed. *Social Media and Democracy*, Cambridge, Cambridge University Press, p. 72

[703] De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union, p. 157

[704] EDRi, 18 November 2019, *Contribution to the examination of France's draft law aimed at combating hate content on the internet*, p. 1

the balancing between the right to free speech and the need to prohibit hate speech is hard to achieve. The emblematic Italian case represented a framework of regulation that still lacks guarantees and protection for discrimination on the basis of sex and gender,[705] while the American case provides a clear and unequivocable example of how some traditions do not view the limitations to freedom of speech as legitimate.[706]

To conclude, the analysis carried out in this work has emphasized the future necessity for a more coherent framework of regulation, starting from the international level, in order to ensure consistency in the application, to avoid abuses by states under the guise of hate speech legislation, and to guarantee that the rights of the individuals and groups targeted by hate speech are respected and protected. Moreover, the study underlined that to ensure an effective combating of the phenomenon, some specific issues should be addressed, i.e., the online dimension of hate speech and the effective application of hate speech norms to all groups protected by anti-discrimination laws.

[705] Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign, p. 22

[706] Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]

Bibliography:

- Abel, R. 1996, *La parola e il Rispetto*, Milano, Giuffrè Editore
- Alaburić, Vesna. 2018. "Legal Concept of Hate Speech and Jurisprudence of the European Court of Human Rights" *Croatian Political Science Review*, Vol. 55, No. 4, pp. 230-252.
- Alkiviadou, Natalie. 2018. "The Legal Regulation of Hate Speech:The International and European Frameworks" *Politicka Misao*, Vol. 4, pp. 203 – 229.
- Amalfitano, C. 2018, *General Principles of EU Law and the Protection of Fundamental Rights*, Cheltenham, Edward Elgar Publishing
- Assimakopoulos, Stavros. Baider, Fabienne. Millar, Sharon. 2017, *Online Hate Speech in the European Union: A Discourse-Analytic Perspective,* Cham: Springer Open.
- Banks, James. Regulating hate speech online, *International Review of Law, Computer and Technology*, Vol. 24:3, pp. 233 – 239.
- Belavusau, Ulad, et al. *Country Case Studies*. Edited by Rumyana Grozdanova and Maria Sperling, International Centre for Counter-Terrorism, 2019, *A Comparative Research Study on Radical and Extremist (Hate) Speakers in European Member States*
- Belavusau, Uladzislau. 2012. "Fighting Hate Speech Through EU Law", *Amsterdam Law Forum*, Vol. 4, no. 1, pp. 20-35
- Benesch, Susan, 2008, "Vile Crime or Inalienable Right: Defining Incitement to Genocide", *Virginia Jurnal of International Law*, Vol 48:3, pp. 485 – 528.
- Berger, J. M., Perez, Heather. 2016, "The Islamic State's Diminishing Returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters", *Occasional Paper GW Program on Extremism*.
- Bertoni, Eduardo. 2005, "Hate Speech under the American Convention on Human Rights", *ILSA Journal of International and Comparative Law*, vol. 12, pp. 569-574.
- Bird, Karen L., "Racist Speech or Free Speech? A Comparison of the Law in France and the United States", *Comparative Politics*, July 2000, Vol. 32, No. 4, pp. 399-418
- Blackstone, W. 1979, *Commentaries on the Laws of England: A Facsimile of the First Edition of 1765--1769*. Chicago, University of Chicago Press.
- Bleich, Erik. July 2011, "The Rise of Hate Speech and Hate Crime Laws in Liberal Democracies", *Journal of Ethnic and Migration Studies*, vol. 37, no. 6, pp. 917 – 934
- Boch, Anna. 2020, "Increasing American Political Tolerance: a framework excluding hate speech", *Socius: Sociological Research for a Dynamic World*, Vol. 6, pp. 1 – 12
- Brattberg, E. 2020, *Reinventing Transatlantic Relations on Climate, Democracy and Technology*, Washington DC, Carnegie Endowment for International Peace
- Brems A., Timmer A., 2016, *Stereotypes and human rights law*, Cambridge: Intersentia
- Brems, E. 2013, *Diversity and European Human Rights – Rewriting judgements of the ECHR*, Cambridge, Cambridge University Press
- Brink, O. David. 2001, "Millian Principles, Freedom of Expression, and Hate Speech", *Legal Theory,* Vol. 7, p. 119 – 157
- Brown, Alexander. "What Is Hate Speech? Part 1: The Myth Of Hate." *Law and Philosophy*, vol. 36, no. 4, 2017, pp. 419–468
- BROWN, ALEXANDER. "WHAT IS HATE SPEECH? PART 2: FAMILY RESEMBLANCES." *Law and Philosophy*, vol. 36, no. 5, 2017, pp. 561–613
- Brown, Alexander. 2015. *Hate Speech Law: a Philosophycal Examination.* New York: Routledge.
- Brown, Alexander. 2018, "What is so special about online (as compared to offline) hate speech?", *Ethnicities*, vol. 18 n. 3, pp. 297 - 326
- Brugger, Winfried. 2003, "The treatment of Hate Speech in German Constitutional Law (part 1)", *German Law Jurnal*, Vol 4. No. 1, pp. 1 – 44

- Burwell, Frances G. *THE CASE FOR ENGAGING EUROPE*. Atlantic Council, 2018, pp. 3–12, *Making America First in the Digital Economy: The Case for Engaging Europe*
- Cacas, Samuel R., and Kathryn A. Tongue. "Countering Cyberhate: More Regulation or More Speech?" *Human Rights*, vol. 25, no. 3, 1998, pp. 21–23
- Campbell, Jude. 2018, "What did the First Amendment originally mean?", *Richmond Law Magazine*, available at: https://lawmagazine.richmond.edu/features/article/-/15500/what-did-the-first-amendment-originally-mean.html
- Carlson, C. R. 2021, *Hate Speech*, Cambridge, The MIT Press
- Carrasco, Luis. 2012, "Redes De Anonimización En Internet: Cómo Funcionan Y Cuáles Son Sus Límites", *Istituto espanol de estudios estrategicos*.
- Casarosa, F. *Handbook on Techniques of Judicial Interaction in the Application of the EU Charter – Freedom of Expression and Countering Hate Speech,* European University Institute, Robert Shuman Centre.
- Castellaneta, Marina. 2017. "Diritti civili e politici – La Corte europea dei diritti umani e l'applicazione del principio dell'abuso del diritto nei casi di hate speech". *Diritti umani e diritto internazionale,* vol. 11, n. 3, pp. 745-751.
- Cheruan, George. 2015, "Hate Speech Law and Policy", *The International Encyclopedia of Digital Communication and Society.*
- Chirco, S. Gori, L. Esposito, I. 2020, "Quando l'odio diventa reato – Caratteristiche e normative di contrasto agli hate crimes", *Poliziamoderna*
- Clucas, Tom. "'Don't Feed the Trolls': Social Media and the Limits of Free Speech." *Violence and Trolling on Social Media: History, Affect, and Effects of Online Vitriol*, edited by Sara Polak and Daniel Trottier, Amsterdam University Press, Amsterdam, 2020, pp. 47–64
- Cogan, N. H. 2015, *The Complete Bill of Rights: The Drafts, Debates, Sources, and Origins*, II ed., New York, Oxford University Press
- Cohen-Almagor, R. 2015, *Confronting the Internet's Dark Side – Moral and social responsibility on the free highway*, New York, Cambridge University Press
- Cohen-Almagor, Raphael. "Freedom of Expression, Internet Responsibility, and Business Ethics: The Yahoo! Saga and Its Implications." *Journal of Business Ethics*, vol. 106, no. 3, 2012, pp. 353–365
- Cohen, Roni. (2014) "Regulating Hate Speech: Nothing Customary about It" *Chicago Journal of International Law*: Vol. 15: No. 1, Article 11, pp. 229 - 255
- Csonka, Peter. 2006, "The Council of Europe's convention on cybercrime and other European initiatives", *Revue Internationale de droit penal*, vol. 77, pp. 473 – 501
- Cuccia V., 2007, *Libertà di espressione e identità collettive*, Torino, G. Giappicchelli Editore
- De Vido, S., Sosa, L. 2021, *Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence - A special report*, Brussels, Publications Office of the European Union
- Delgado, R. Stefancic, J. 2018, *Must we defend Nazis? – Why the First Amendment Should not Protect Hate Speech and White Supremacy*, New York, New York University Press
- Duarte, N., Llanso, E., Loup, A. 2017. *Mixed Messages? The limits of automated social media content analysis*, Center for Democracy and Technology
- Echikson, W. Knodt, O., 2018, *Germany's NetzDG: A key test for combatting online hate*, CEPS Digital forum, Counter Extremism Project
- Ed. Keane D., McDermott Y., 2012, *The Challenge of Human Rights Past, Present and Future*, Northampton, Edward Elgar Publishing
- Ed. Langford M., 2008, *Social Rights Jurisprudence*, Cambridge: Cambridge University Press

- EDRi, 18 November 2019, *Contribution to the examination of France's draft law aimed at combating hate content on the internet*.
- Edwards, A. 2011, *Violence Against Women Under International Human Rights Law*, Cambridge, Cambridge University Press
- Elthaway, Mona. "WHY DO THEY HATE US?" *Foreign Policy*, no. 193, 2012, pp. 64–84.
- Farrior, Stephanie. 1996. "Molding the Matrix: The Historiacl and Theoretical Foundations of International Law Concerning Hate Speech". *Berkeley Jurnal of International Law,* vol 14: No. 1, pp. 1-98.
- Feldman, Stephen M. Hate Speech and Democracy, 2013, Criminal Justice Ethics, Vol.32, no. 1, pp. 78-90
- Fisch, William B. "Hate Speech in the Constitutional Law of the United States", *The American Journal of Comparative Law*, 2002, Vol. 50, pp. 463-492
- Fredman S., 2011, *Discrimination Law*, Oxford: Oxford university press
- Fredman, Sandra. 2018. "Freedom of speech" in *Comparative Human Rights Law,* pp. 305 - 354. Oxford: Oxford University Press.
- Fried, C., 1992. The New First Amendment Jurisprudence: A Threat to Liberty. *The University of Chicago Law Review*, 59(1), p.225.
- Frigo, M. Lang, A. 2005, *Diritto dell'Unione Europea e della Comunità Internazionale – Casi e Materiali*, Torino, Giappichelli Editore
- Futtner, N. Brusco, N. 2021, *Hate Speech is on the Rise – Report*, Geneva, Geneva International Centre for Justice
- Ganesh, Bharath. 2018, "The Ungovernability Of Digital Hate Culture." *Journal of International Affairs*, vol. 71, no. 2, pp. 30–49
- Gelber, K. 2002, *Speaking back : the free speech versus hate speech debate*, Amsterdam, John Benjamins Publishing
- Ghanea, Nazila. "Intersectionality and the Spectrum of Racist Hate Speech: Proposals to the UN Committee on the Elimination of Racial Discrimination." *Human Rights Quarterly*, vol. 35, no. 4, 2013, pp. 935–954
- Giusti, G. Iannàccaro, G. 2020, Language, Gender and Hate Speech – A Multidisciplinary Approach, Venezia, Edizioni Ca' Foscari
- Gledhill, K. 2015, *Human Rights Acts – The Mechanisms Compared*, Oxford, Hart Publishing
- Global Freedom of Expression. 2021. *Sürek and Özdemir v. Turkey - Global Freedom of Expression*. [online] Available at: <https://globalfreedomofexpression.columbia.edu/cases/surek-ozdemir-v-turkey/> [Accessed 13 November 2021].
- Goldberg, Erica. "FREE SPEECH CONSEQUENTIALISM." *Columbia Law Review*, vol. 116, no. 3, 2016, pp. 687–756
- Goldberg, Erica. "Free Speech Consequentialism." *Columbia Law Review*, vol. 116, no. 3, 2016, pp. 687–756.
- Gometz, Gianmarco. 2017, "L'odio proibito: la repression giuridica dello Hate Speech", Rivista Telematica www.statoechiese.it, n. 32
- Gomez Isa, Felipe. 2003, "The Optional Protocol For The Convention On The Elimination Of All Forms Of Discrimination Against Women: Strengthening The Protection Mechanisms Of Women's Human Rights", *Arizona Journal of International and Comparative Law*, vol. 20, no. 2, pp. 291 – 321
- Green, Donald P., et al. "Hate Crime: An Emergent Research Agenda." *Annual Review of Sociology*, vol. 27, 2001, pp. 479–504
- Heinze, E. 2016, *Hate Speech and Democratic Citizenship*, Oxford, Oxford University Press

- Heinze, Eric. “Viewpoint Absolutism and Hate Speech.” *The Modern Law Review*, vol. 69, no. 4, 2006, pp. 543–582.
- Heinze, Erik. 2013, “Review Essay - Hate Speech and the normative foundations of regulation”, International Jurnal of Law in Context, Vol 9, pp. 590 – 617
- Heldt, Amélie. “Let's Meet Halfway: Sharing New Responsibilities in a Digital Age.” *Journal of Information Policy*, vol. 9, 2019, pp. 336–369
- Henry, S. Jessica, 2009, “Beyond free speech: novel approaches to hate on the Internet in the United States, *Information and communications Technology law*, vol. 18, no. 2, pp. 235 – 251
- Hepple, B. Szyszczak, E. M. 1993, *Discrimination: The Limits of Law*, London, Mansell Publishing
- Herz, M. Molnar, P. 2012, *The Content and Context of Hate Speech – Rethinking Regulation and Responses*, Cambridge, Cambridge University Press
- Howard, E. 2018*, Freedom of Expression and Religious Hate Speech in Europe*, New York, Routledge
- Janssen, E. "Faith in Public Debate. On Freedom of Expression”,2015, *School of Human Rights Research Series*, no. 68
- John T. Benne, *the Harm in Hate Speech: A Critique of the Empirical and Legal Bases of Hate Speech Regulation*, 43 Hastings Const. L.Q. 445 (2016).
- Johnson, A. 2001, *Privilege, Powe and Difference*, New York, McGraw-Hill.
- Kang, M. Rivé-Lasan, M. Kim, W. Hall, P. 2020, *Hate Speech in Asia and Europe – Beyond Hate and Fear*, New York, Routledge.
- Keats Citron, D. 2014, *Hate Crimes in Cyberspace*, Cambridge, London, Harvard University Press
- Kemekenidou, Penelope. “r/ChokeABitch: Feminist Tactics Against Hate Speech in Capitalist Social Media Platforms.” *Violence and Trolling on Social Media: History, Affect, and Effects of Online Vitriol*, edited by Sara Polak and Daniel Trottier, Amsterdam University Press, Amsterdam, 2020, pp. 233–250
- Kendell, Kate. “Ending Discrimination Based on Sexual Orientatíon and Gender Identity.” *Human Rights*, vol. 35, no. 4, 2008, pp. 18–19
- Khaitan, T. 2015, *A Theory of Discrimination Law*, Oxford, Oxford University Press.
- Lane Burner, M., Balter-Reitz, Susan. 2013, “Snyder v. Phelps: The U.S. Supreme Court's Spectacular Erasure of the Tragic Spectacle”, *Rethoric and Public Affairs*, Vol. 16, no. 4, pp. 651 - 683
- Langton, Rae. 1993, “Speech Acts and Unspeakable Acts”, *Philosophy and Public Affairs*,Vol. 22, No. 4. pp. 293-330.
- Langton, Rae. 2018 “The Authority of Hate Speech”, *Oxford Studies in Philosophy of Law*, Vol. 3.
- Latraverse, S. 2011 *Report on measures to combat discrimination – Directives 2000/43/EC and 2000/78/EC*, Country report France, European Network of Legal Experts in the Non-discrimination Field
- Le Quadrature du Net, 27 juin 2019, *Letter to the parliament - PPL Avia, refusez les mesures inutiles et dangereuses*
- Lepoutre, Maxime. “Hate Speech in Public Discourse: A Pessimistic Defense of Counterspeech.” *Social Theory and Practice*, vol. 43, no. 4, 2017, pp. 851–883.
- Levin, A. 2010,*The Cost of Free Speech – Pornography, Hate Speech and their Challenge to Liberalism,* New York, Palgrave Macmillan
- Levin, Abigail. 2009, “Pornography, Hate Speech, and Their Challenge to Dworkin's Egalitarian Liberalism”, *Public Affairs Quarterly*, Vol. 23 No. 4, pp. 357-373

- Lokiec, Pascal. "Discrimination Law in France", *University Paris XIII*, Available at: https://www.jil.go.jp/english/events/documents/clls08_lokiec.pdf
- Marks, S., 2006. Global Recognition of Human Rights for Lesbian, Gay, Bisexual, and Transgender People. *Health and Human Rights*, 9(1), p.33.
- Mchangama, J. 2015, "THE PROBLEM WITH HATE SPEECH LAWS", *The Review of Faith & International Affairs*, vol. 13 , pp. 75 - 82
- McPHAIL, BEVERLY A. "GENDER-BIAS HATE CRIMES: A Review." *Trauma, Violence & Abuse*, vol. 3, no. 2, 2002, pp. 125–143.
- Meyer, Doug. "Evaluating the Severity of Hate-Motivated Violence: Intersectional Differences among LGBT Hate Crime Victims." *Sociology*, vol. 44, no. 5, 2010, pp. 980–995.
- Meyer, Doug. "Interpreting and Experiencing Anti-Queer Violence: Race, Class, and Gender Differences among LGBT Hate Crime Victims." *Race, Gender & Class*, vol. 15, no. 3/4, 2008, pp. 262–282
- Micklitz, H. W. Tridimas, T. 2015, *Risks and EU law*, Cheltenham, Edward Elgar Publishing
- Molnar, P. Ed. 2015, *Free Speech and Censorship Around the Globe*, Budapest, Central European University Press.
- Murphy, S., 2002. Hate-Speech Protocol to Cybercrime Convention. *American Journal of International Law*, 96(4), pp.973-975.
- Open Society Foundations. *ANNEX I: SPECIFIC RECOMMENDATIONS*. Open Society Foundations, 2017, pp. 27–32, *REVISITING THE EU ROMA FRAMEWORK: Assessing the European Dimension for the Post-2020 Future*
- Orofino, M. 2014, *La libertà di espressione tra costituzione e carte europee dei diritti – Il dinamismo di una società in continua trasformazione*, Torino, Giappichelli Editore
- Pálmadóttir, Jóna Aðalheiður. Kalenikova, Iuliana. 2018. "*Hate speech: an overview and reccommendations for combating i*t", Reykjavik: Icelandic Human Rights Centre.
- Pamment, James. *Current EU Policy on Disinformation*. Carnegie Endowment for International Peace, 2020, pp. 5–10, *The EU's Role in Fighting Disinformation: Taking Back the Initiative*
- Phelps-Roper, M. 2019, *Unfollow A memoir of loving and leaving the Westboro Baptist Church*, New York, Farrar, Straus and Giroux
- Phillips, S. and Grattet, R., 2000. Judicial Rhetoric, Meaning-Making, and the Institutionalization of Hate Crime Law. *Law & Society Review*, 34(3), p.567.
- Pitch T., 2004, *I diritti fondamentali: differenze culturali, disuguaglianze sociali, differenza sessuale*, Torino, G. Giappicchelli Editore
- Poland, B. 2016, *Haters harassment, abuse and violence online*, University of Nebraska, Potomac Books
- Posenato, Naiara, 2015, "The protection of the right to freedom of expression: A panorama of the Inter-American Court of Human Rights case law", *Espaço Jurídico Journal of Law*, Vol. 16: No. 3, p. 51-68.
- Posenato, Naiara. "The Protection Of The Right To Freedom Of Expression: A Panorama Of The Inter-American Court Of Human Rights Case Law", Espaço Juridico Jurnal of Law (EJJL), v. 16 (2015), Edição Especial - Graves Violações DDHH. pp. 51 – 68.
- Richardson-Self, L. 2021, *Hate Speech against Women Online - Concepts and Countermeasures*, London, Rowman and Littlefield
- Rosa, Alessandro di. 2019. "Performative Hate Speech Acts." *The Age of Human Rights Journal*, June 12: 105-132.
- Rose, Hadley. "SPEAK NO EVIL, HEAR NO EVIL, DO NO EVIL: HOW RATIONALES FOR THE CRIMINALIZATION OF HATE SPEECH APPLY IN TRANSITIONAL CONTEXTS." *Willamette Journal of International Law and Dispute Resolution*, vol. 22, no. 2, 2015, pp. 313–342

- Rosenfeld, Michel. 2003, "Hate Speech in Constitutional Jurisprudence: A Comparative Analysis ", *Cardozo Law Review*, pp. 1523 - 1567.
- Rubenfeld, J., 2001. The First Amendment's Purpose. *Stanford Law Review*, vol. 53, pp. 770 - 832.
- Sabrina Alimahomed-Wilson. "Invisible Violence: Gender, Islamophobia, and the Hidden Assault on U.S. Muslim Women." *Women, Gender, and Families of Color*, vol. 5, no. 1, 2017, pp. 73–97
- Scaffardi, L. 2018, *I "profili" del diritto – Regole, rischi e opportunità nell'era digitale*, Torino, Giappichelli Editore
- Scrivens, Ryan. "115 Academic Theses (Ph.D. and MA) on the Role of the Internet in Facilitating and Combating Radicalization, Extremism, Terrorism and Cyber-Terrorism, Written in English between 1995 and 2019." *Perspectives on Terrorism*, vol. 13, no. 3, 2019, pp. 230–237
- Siegel, A. 2020, *Online Hate Speech,* in Persily, N. Tucker, J ed. *Social Media and Democracy*, Cambridge, Cambridge University Press
- SIMPSON, ROBERT MARK. "DIGNITY, HARM, AND HATE SPEECH." *Law and Philosophy*, vol. 32, no. 6, 2013, pp. 701–728.
- Softness, Nicole. "Terrorist Communications: Are Facebook, Twitter, and Google Responsible for the Islamic State's Actions?" *Journal of International Affairs*, vol. 70, no. 1, 2016, pp. 201–215
- Sorial, Sarah. "Hate Speech And Distorted Communication: Rethinking The Limits Of Incitement." *Law and Philosophy*, vol. 34, no. 3, 2015, pp. 299–324.
- Steiner, H. J. Alston, P. Goodman, R. 2007, *International Human Rights Law in Context*, Oxford, Oxford University Press
- Stone, Geoffrey. 1994, "Hate Speech and the U.S. Constitution", *East European Constitutional Review*, Vol. 78, pp. 78 - 82.
- Stradella, E. 2008, *La libertà di espressione politico-simbolica e I suoi limiti: tra teorie e "prassi"*, Torino, Giappichelli Editore
- Strossen, N. 2018, *HATE - Why We Should Resist It with Free Speech, Not Censorship,* Oxford, Oxford University Press
- Synodinou, T. E. Jougleux, P. Markou, C. prastitou, T. 2017, *EU Internet Law*, Cham, Springer International Publishing
- Tambini, Damian, and Centre for International Governance Innovation. *Rights and Responsibilities of Internet Intermediaries in Europe: The Need for Policy Coordination*. Centre for International Governance Innovation, 2019, pp. 91–95
- Tanzarella, P. 2020, *Discriminare parlando – Il pluralism democratico messo alla prova dai discorsi d'odio razziale*, Torino, Giappichelli Editore
- Traum, Alexander. "Contextualising the Hate Speech Debate: the United States and South Africa." *The Comparative and International Law Journal of Southern Africa*, vol. 47, no. 1, 2014
- Tworek, H. Leerssen, P., 2019, *An Analysis of Germany's NetzDG Law*, Transatlantic Working Group
- Úbeda de Torres, Amaya. "Freedom of Expression under the European Convention on Human Rights: A Comparison With the Inter-American System of Protection of Human Rights." *Human Rights Brief* 10, no. 2 (2003): p. 6-12.
- Ülgen, Sinan. *FREEDOM OF EXPRESSION ONLINE*. Carnegie Endowment for International Peace, 2016, pp. 17–28, *GOVERNING CYBERSPACE: A Road Map for Transatlantic Leadership*
- Van Der Wilk, Adriane. 2018, "Cyber violence and hate speech online against women", *Research Paper for the European Parliament's Committee on Women's Rights and Gender Equality*, European Union.

- Van Dijk, P. Van Hoof, F. Van Rijn, A. Zwaak, L. 2018, *Theory and practice of the European Convention on Human Rights,* Cambridge, Intersentia
- Vitucci M. C., 2012, *La tutela Internazionale dell'orientamento sessuale*, Napoli, Jovene Editore
- Waldron, J. 2012, *The Harm In Hate Speech*, Cambridge, London, Harvard University Press.
- Weber, A. 2009, *Manual on Hate Speech*, Strasbourg, Council of Europe Publishing.
- Wendel, W. Bradley. 2002, "Certain Fundamental Truths: A Dialectic on Negative and Positive Liberty in Hate Speech Cases", *Law and Contemporary Problems*, Vol. 65, No. 2, pp. 33-85
- Weston-Scheuber, Kylie. 2012, "Gender and the Prohibition of Hate Speech", *Law and Justice Jurnal*, vol. 12, no. 2, pp. 132 - 150
- Wojciezak, Magdalena. 2010, "'Don't talk to me': effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism", *New media and society Sage*, Vol 12 No. 4, pp. 637 - 655
- Yong, Caleb. 2011, "Does freedom of speech include hate speech?", *Res Publica*, vol. 17, no. 4, pp. 385 – 403.
- Zagrebelsky, V. Chenal, R. Tomasi, L. 2019, *Manuale dei Diritti fondamentali in Europa*, Bologna, Il Mulino
- Ziccardi, G. 2020, *Online Political Hate speech in Europe*, Cheltenham, Edward Elgar Publishing

Cases and Documents:

- *ADVISORY OPINION OC-25/18 OF 30 MAY 2018 REQUESTED BY THE REPUBLIC OF ECUADOR*, Inter-American Court of Human Rights (IACrtHR), 30 May 2018
- Article 19, 2017, *Germany: the Act to Improve Enforcement of the Law in Social Networks*, Media Against Hate Campaign
- Article 19, 2018, *Italy: responding to "hate speech"*, Media Against Hate Campaign
- Article 19, 2018, *Responding to "hate speech": comparative overview of six EU countries*, Media Against Hate Campaign
- Beizaras And Levickas V. Lithuania, App. No. 41288/15 (ECtHR, 14 January 2020)
- Belkacem v. Belgium, App. No. 34367/14 (ECtHR, 27 June 2017).
- BVerfGE 90, 241, 247, Decision of 13 April 1994, Auschwitz Lie Case (Holocaust Denial Case) = Decisions 620, at 625.
- Carl Jóhann Lilliendahl v. Iceland, App. No. 29297/18 (ECtHR, 11 June 2020)
- Cass., sez. III, 23 giugno 2015, n. 2742/2015, Giur. It. 2015
- Cass., sez. III, 24 aprile 2013, n. 1126/2013, Giur. It. 2013
- Committee on the Elimination of Racial Discrimination*, General Recommendation VII relating to the implementation of Article 4 of the Convention, (Thirty-second session, 1985), Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies,U.N. Doc. HRI\GEN\1\Rev.1 at 64 (1994).*. [online] Available at: http://hrlibrary.umn.edu/gencomm/genrevii.htm
- Conseil constitutionnel [CC] [Constitutional Court] decision No. 2020-801 DC, June 18, 2020, Rec. France
- Council of Europe Committee of Ministers, Recommandation No. R (97) 21 of the Committee of Ministers to member states on the Media and the promotion of a culture of Tolerance (Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies)

- Council of Europe Committee of Ministers, Recommandation No. R (97) 21 of the Committee of Ministers to member states on “Hate Speech” (Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies)
- Council of Europe Committee of Ministers, *Recommendation CM/Rec(2019)1 of the Committee of Ministers to member States on preventing and combating sexism* (*Adopted by the Committee of Ministers on 27 March 2019 at the 1342nd meeting of the Ministers' Deputies)*
- Council of Europe Gender Equality Unit, *Background note on Sexist Hate Speech*, 1 February 2016
- Council of Europe, Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems, European Treaty Series, No. 189, Strasbourg, 28.I.2003.
- Council of Europe, *Council of Europe Convention on preventing and combating violence against women and domestic violence*, 11 May 2011, available at: https://www.coe.int/en/web/istanbul-convention/about-the-convention
- Council of Europe, ELSA, *International Legal Research Group on ONLINE HATE SPEECH – Final Report*, (February 2014)
- Council of Europe, European Commission against Racism and Intolerance (ECRI) General Policy Recommendation No. 15 on combating hate speech adopted on 8 december 2015.
- Council of Europe, March 2018, *Gender equality strategy 2018-2023,* (Adopted by the Committee of Ministers of the council of Europe)
- Council of the European Union, *Charter of Fundamental Rights of the European Union (2007/C 303/01)*, 14 December 2007, C 303/1, Available at: https://www.europarl.europa.eu/charter/pdf/text_en.pdf
- Council of the European Union, FRAMEWORK DECISION 2008/913/JHA, on combating certain forms and expressions of racism and xenophobia by means of criminal law, L 328/55
- Council of the European Union, Joint action 96/443/JHA, concerning action to combat racism and xenophobia, OJ L 185, 24.7.96, *EU Anti-Discrimination Policy: Annexes,* Available at: https://www.europarl.europa.eu/workingpapers/libe/102/text5_en.htm#annex4
- Decreto Legislativo 9 July 2003, n. 215, in G. U. Aug. 12, 2003, n. 186
- Directorate-General for Internal Policies, *The European legal framework on hate speech, blasphemy and its interaction with freedom of expression*, 2015
- Erbakan v. Turkey, App. No. 59405/00 (ECtHR, 6 July 2006)
- European Commission, Factsheet – 6th evaluation of the Code of Conduct, Directorate-General for Justice and Consumers, 7 October 2021.
- European Commission, March 2020, *A Union of Equality: Gender Equality Strategy 2020-2025*, Brussels
- European Commission, PRISM - Preventing Redressing & Inhibiting hate Speech in new Media, *Hate Crime and Hate Speech in Europe: Comprehensive Analysis of International Law Principles, EU-wide Study and National Assessments*, 2015
- European Court of Human Rights, *Factsheet - Hate speech*, Press Unit, June 2020.
- European Court of Human Rights, *Guide on Article 17 of the European Convention on Human Rights,* 8/31/2019.
- European Parliament and Council Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), L 178/1
- European Parliament and Council Directive 2015/1535 of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (codification), L 241/1

- European Parliament resolution, 2019/2933(RSP) of 18 December 2019 on public discrimination and hate speech against LGBTI people, including LGBTI free zones, P9_TA(2019)0101.
- European Parliament resolution, B4-0731/95 of 27 April 1995, Resolution on racism, xenophobia and anti-Semitism, Official Journal C 126, 22/05/1995 P. 0075.
- Féret v. Belgium, App. No. 15615/07 (ECtHR, 16 July 2009)
- French 1946 Const. Preamble, France
- French 2008 Const., France
- Garaudy v. France, App. no. 65831/01 (ECtHR, 7 July 2003)
- General recommendation No. 35, Committee on the Elimination of Racial Discrimination, 26 September 2013
- Ghanea, Nazila, 2012, "*The Concept of Hate speech and its Evolution overtime*", United Nations Committee on the Elimination of Racial Discrimination's day of thematic discussion on Racist Hate Speech, Geneva.
- Gündüz v. Turkey, App. No. 35071/97 (ECtHR, 4 December 2003)
- Handyside V. The United Kingdom, App. no. 5493/72 (ECtHR, 7 December 1976)
- Hogan Lovells, *The Global Regulation of Online Hate: A Survey of Applicable Laws – Special Report*, Hogan Lovells for PeaceTech Lab (December 2020)
- Italian 1947 Const., Italy
- Ivanov v. Russia, App No. 35222/04 (ECtHR, 20 February 2007)
- Jersild V. Denmark, App. No. 15890/89 (ECtHR, 23 September 1994)
- Legge 20 June 1952, n. 645, in G. U. June, 23, 1952, n. 143
- Legge 22 May 1975, n. 152, in G. U. May, 24, 1975, n. 136
- Legge 25 June 1993, n. 205, in G. U. June, 26, 1993, n. 148
- Legge 3 May 2004, n. 112, in G. U. May. 5, 2004, n. 82
- Lilliendhal v. Iceland, App. No. 29297/18 (ECtHR, 12 May 2020)
- M'Bala M'Bala v. France, App. No. 25239/13 (ECtHR, 10 November 2015)
- Marais v. France, App. No. 31159/96 (ECtHR, 24 June 1996)
- McGonagle, Tarlach. 2013. "*The Council of Europe against online hate speech: Conundrums and Challenges*". Republic of Serbia, Ministry of Culture and Information, Expert paper for the Council of Europe Conference of Ministers responsible for Media and Information Society, 'Freedom of Expression and Democracy in the Digital Age: Opportunities, Rights, Responsibilities', Belgrade, 7-8 November.
- Ministers of the Freedom Online Coalition, *Reccomendations for Freedom Online*, Tallin, April 28, 2014
- Ministry of Foreign Affairs and International Cooperation, Inter-ministerial Committee for Human Rights, *Italy's Remarks following communication from un Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on the red button protocol*, Mr. David Kaye (May 2018)
- Norwood v. The United Kingdom, App. No. 23131/03 (ECtHR, 16 November 2004).
- OAS, IACHR, *Annual of the office of the Special Rapporteur for Freedom of Expression*, Office of the Special Rapporteur for Freedom of Expression, Official records; OEA/Ser.L/V/II, (2015).
- OAS, IACHR, *Violence Against LGBTI Persons*, Official records; OEA/Ser.L, 2015
- Office on Genocide Prevention and the Responsibility to Protect, *United Nations Strategy and Plan of Action on Hate Speech – Detailed Guidance on Implementation for United nations Field Presence*, United Nations, 2020, available at: https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml

- Organization of American States, *Interamerican Convention against all forms of discrimination and intolerance*, June 5, 2013, A-69, available at: http://www.oas.org/en/sla/dil/inter_american_treaties_A-69_discrimination_intolerance.asp
- Organization of American States, *Interamerican Convention against racism, racial discrimination and related forms of intolerance*, June 5, 2013, A-68, available at: http://www.oas.org/en/sla/dil/inter_american_treaties_A-68_racism.asp
- Organization of American States. 2013. *Inter-American Convention Against All Forms of Discrimination and Intolerance*. Organization of American States.
- Pastörs v. Germany, App. No. 55225/14 (ECtHR, 3 October 2019)
- Perinçek V. Switzerland, App. No. 27510/08 (ECtHR, 15 October 2015)
- Proposition de loi 20 March 2019, n. 1785, *visant à* lutter contre *la* haine *sur* internet, Assemblée Nationale
- Proposition de loi 22 Janvier 2020 n. 388, *visant à* lutter contre *la* haine *sur* internet, Assemblée Nationale
- Roj TV A/S v Denmark, App. No. 24683/14 (ECtHR, 17 April 2018)
- Schenck v. U.S., 249 U.S. 47 (1919)
- Senato della Repubblica, *Disegno di Legge (Misure di prevenzione e contrasto della discriminazione e della violenza per motivi fondati sul sesso, sul genere, sull'orientamento sessuale, sull'identità di genere e sulla disabilità)*, n. 2005, XVIII legislatura, 5 November 2020
- Snyder v. Phelps, App. No. 09 – 751 (SCOTUS, 2 March 2011)
- Soulas v. France, App. No. 15948/03 (ECtHR, 10 October 2008)
- Tagiyev And Huseynov V. Azerbaijan, App. No. 13274/08 (ECtHR, 5 December 2019)
- The Prosecutor V. Ferdinand Nahimana Jean-Bosco Barayagwiza Hassan Ngeze, Case No. ICTR-99-52-T (ICTR, 3 December 2003)
- U.S. 1789 Const. Amend. 1, U.S.
- UN Committee on the Elimination of Racial Discrimination (CERD), *General recommendation No. 35 : Combating racist hate speech*, 26 September 2013, CERD/C/GC/35
- UN General Assembly, *Convention on the Elimination of All Forms of Discrimination Against Women*, 18 December 1979, United Nations, Treaty Series, vol. 1249, p. 13, available at: https://www.ohchr.org/EN/ProfessionalInterest/Pages/CEDAW.aspx
- UN General Assembly, *International Convention on the Elimination of All Forms of Racial Discrimination*, 21 December 1965, United Nations, Treaty Series, vol. 660, p. 195, available at: https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx.
- UN General Assembly, *International Covenant on Civil and Political Rights*, 16 December 1966, United Nations, Treaty Series, vol. 999, p. 171, available at: https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx.
- UN General Assembly. (1948). *Universal declaration of human rights* (217 [III] A). Paris. Available at: https://www.un.org/en/about-us/universal-declaration-of-human-rights.
- UN Human Rights Committee (HRC), *CCPR General Comment No. 11: Article 20 Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred*, 29 July 1983, available at: https://www.ohchr.org/Documents/Issues/Opinion/CCPRGeneralCommentNo11.pdf
- UN Human Rights Committee (HRC), *General comment no. 34, Article 19, Freedoms of opinion and expression*, 12 September 2011, CCPR/C/GC/34, available at: https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
- UN Human Rights Council, *Annual report of the United Nations High Commissioner for Human Rights : Addendum, 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, 11 January 2013, A/HRC/22/17/Add.4, available at: https://www.ohchr.org/en/issues/freedomopinion/articles19-20/pages/index.aspx
- UN Human Rights Council, Resolution on Human rights, sexual orientation and gender identity, 14 July 2011, A/HRC/RES/17/19
- UN Human Rights Council, *Resolution on The promotion, protection and enjoyment of human rights on the Internet*, 1 July 2016, A/HRC/RES/32/13
- UNAR, Ufficio per la promozione della parità di trattamento e la rimozione delle discrminazioni fondate sulla razza o sull'origine etnica, *Relazione al parlamento sull'attività svolta e sull'effettiva applicazione del principio di parità di trattamento e sull'efficacia dei meccanismi di tutela*, 2020.
- UNGA, *Annual report of the United Nations High Commissioner for Human Rights*, UNGAOR, 22nd sess, UN Doc A/HRC/22/17/Add.4 (2013)
- UNGA*, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UNGAOR, 74th Sess, Supp No. 486, UN Doc A/74/486 (2019).
- UNGA, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/74/486 (9 October 2019)
- UNGA*, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/67/357 (7 September 2012)
- UNGA, *Report of the Special Rapporteur on violence against women, its causes and consequences*, A/HRC/17/26 (2 May 2011)
- United States Court of Appeal for the Ninth Circuit, Mehier Taamneh; Lawrence Taamneh; Sara Taamneh; Dimana Taamneh, V. Twitter, Inc.; Google Llc; Facebook, Inc., No. 18-17192, D.C. No. 3:17-cv-04107-EMC
- Vedjdeland and Others v. Sweden, App. No. 1813/07 (ECtHR, 9 February 2012)
- Venice Commission, 2010. *Blasphemy, insult and hatred: finding answers in a democratic society*, Science and technique of democracy, No. 47, Council of Europe Publishing
- Williamson v. Germany, App. No. 64496/17 (ECtHR, 31 January 2019).
- Yahoo! Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, 169 F. Supp. 2d at

Sitography:

- Article 19 website: https://www.article19.org/about-us/ [Accessed 23 January 2022]
- Article 19 website: https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/). [Accessed 23 January 2022]
- Article 19 website: https://www.article19.org/resources/italy-responding-hate-speech/ [Accessed 20 January 2022]
- Associazione per gli Studi Giuridici dull'Immigrazione website: http://www.asgi.it/banca-dati/corte-dappello-trento-sezione-penale-sentenza-del-1-giugno-2016/ [Accessed 22 January 2022]
- BBC website: https://www.bbc.com/news/world-us-canada-45968606 [Accessed 30 January 2022]
- Britannica website: https://www.britannica.com/biography/Matthew-Shepard [Accessed 30 January 2022]
- Britannica website: https://www.britannica.com/event/United-States-Capitol-attack-of-2021 [Accessed 3 February 2022]
- Carabinieri website: http://www.carabinieri.it/arma/partners/osservatorio-per-la-sicurezza-contro-gli-atti-discriminatori-oscad [Accessed 24 January 2022]
- Constitution Annotated website: https://constitution.congress.gov/browse/essay/amdt1_2_1/ [Accessed 28 January 2022]

- Corriere della Sera website: https://www.corriere.it/cronache/21_giugno_22/vaticano-ddl-zan-legge-testo-b13294ba-d2d0-11eb-9207-8df97caf9553.shtml [Accessed 26 January 2022]
- Council of Europe Website: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=185 [Accessed 14 Jaunary 2022]
- Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=189 [Accessed 18 January 2022]
- Council of Europe website: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=185 [Accessed 18 January 2022]
- Diritto.it website: https://www.diritto.it/ddl-zan-analisi-di-una-proposta-di-legge-molto-discussa/ [Accessed 26 January 2022]
- Dutch Government website: https://wetten.overheid.nl/BWBR0001854/2020-07-25/#BoekTweede_TiteldeelV_Artikel137d [Accessed 8 January 2022].
- EDRi website: https://edri.org/about-us/ [Accessed 24 January 2022]
- EDRi website: https://edri.org/our-work/frances-law-on-hate-speech-gets-thumbs-down/ [Accessed 24 January 2022]
- Eur-lex website: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l33178 [Accessed 4 December 2021].
- Eur-lex website: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=IT [Accessed 25 January 2022]
- European Commission website: https://ec.europa.eu/growth/tools-databases/tris/en/index.cfm/search/?trisaction=search.detail&year=2019&num=412&mLang=EN [Accessed 23 January 2022]
- European Commission website: https://ec.europa.eu/growth/tools-databases/tris/en/about-the-20151535/the-notification-procedure-in-brief1/ [Accessed 11 February 2022]
- European Commission website: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en [Accessed 16 January 2022]
- European Commission website: https://ec.europa.eu/newsroom/just/items/31811/en [Accessed 3 December 2021].
- European Parliament website: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2020/0361(COD)&l=en [Accessed 18 January 2022]
- European Parliament website: https://www.europarl.europa.eu/doceo/document/TA-7-2014-0062_EN.html [Accessed 3 December 2021].
- European Parliament website: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act [Accessed 18 January 2022]
- European Parliament website: https://www.europarl.europa.eu/news/it/press-room/20211210IPR19209/digital-services-act-safer-online-space-for-users-stricter-rules-for-platforms [Accessed 18 January 2022]
- European Union Agency for Fundamental Rights website: https://fra.europa.eu/en/data-and-maps/2020/lgbti-survey-data-explorer [Accessed 25 January 2022]
- European Website on Integration website: https://ec.europa.eu/migrant-integration/library-document/hate-crime-and-hate-speech-europe-comprehensive-analysis-international-law_en [Accessed 18 November 2021].
- Facebook community standards website: https://transparency.fb.com/it-it/policies/community-standards/ [Accessed 16 January 2022]
- French Constitutional Council website: https://www.conseil-constitutionnel.fr/en [Accessed 25 January 2022]

- French Government website: https://www.gouvernement.fr/en/everything-you-need-to-know-about-freedom-of-expression-in-france [Accessed 18 January 2022]
- French Senate website: https://www.senat.fr/presse/cp20200619b.html [Accessed 24 January 2022]
- German Federal Ministry of the Justice website: https://www.gesetze-im-internet.de/englisch_gg/ [Accessed 9 January 2022].
- German Law Archive website: https://germanlawarchive.iuscomp.org/?p=1245 [Accessed 5 February 2022]
- Global Freedom of Expression website: https://globalfreedomofexpression.columbia.edu/cases/la-colegiacion-obligatoria-de-periodistas-oc-0585/ [Accessed 8 December 2021].
- History website: https://www.history.com/topics/united-states-constitution/first-amendment#:~:text=The%20First%20Amendment%20to%20the,speech%2C%20religion%20and%20the%20press.&text=The%20amendment%20was%20adopted%20in,civil%20liberties%20under%20U.S.%20law [Accessed 28 January 2022]
- History, Art and Archives United States House of Representatives website: https://history.house.gov/Historical-Highlights/1700s/The-Sedition-Act-of-1798/ [Accessed 28 January 2022]
- Human Rights Watch website: https://www.hrw.org/news/2018/05/30/frances-creeping-terrorism-laws-restricting-free-speech [Accessed 20 January 2022]
- ILGA Europe website: https://www.ilga-europe.org/what-we-do/our-advocacy-work/hate-crime-hate-speech [Accessed 3 February 2022]
- International Justice Resource Center website: https://ijrcenter.org/un-treaty-bodies/committee-on-the-elimination-of-discrimination-against-women/ [Accessed 17 January 2022]
- Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/249/47/#tab-opinion-1928047 [Accessed 29 January 2022]
- Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/250/616/ [Accessed 29 January 2022]
- Justitia US Supreme Court Center website: https://supreme.justia.com/cases/federal/us/391/367/ [Accessed 29 January 2022]
- Liberation website: https://www.liberation.fr/debats/2020/01/21/feministes-lgbti-et-antiracistes-nous-ne-voulons-pas-de-la-loi-cyberhaine_1774297/ [Accessed 24 January 2022]
- Library of Congress website: https://www.loc.gov/item/global-legal-monitor/2020-06-29/france-constitutional-court-strikes-down-key-provisions-of-bill-on-hate-speech/ [Accessed 22 January 2022]
- Organization of American States webiste: https://www.oas.org/en/iachr/mandate/Basics/intro.asp> [Accessed 5 December 2021].
- Organization of American States website: https://www.oas.org/en/iachr/expression/showarticle.asp?artID=443&lID=1https://www.oas.org/en/iachr/expression/showarticle.asp?artID=443&lID=1 [Accessed 5 December 2021].
- Politico website: https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/ [Accessed 4 February 2022]
- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d’Odio website: https://www.retecontrolodio.org/chi-siamo/ [Accessed 23 January 2022]
- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d’Odio website: https://www.retecontrolodio.org/protocollo-d-intesa/ [Accessed 23 January 2022]

- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/01/27/osservazioni-sulle-leggi-contro-la-discriminazione/ [Accessed 25 January 2022]
- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/10/28/ddl-zan-uno-stop-che-brucia-ma-che-non-ci-fermera/ [Accessed 25 January 2022]
- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/03/29/legge-contro-l-omo-bi-transfobia-la-misoginia-e-l-abilismo-a-quando-la-calendarizzazione-in-senato/ [Accessed 26 January 2022]
- Rete Nazionale per il Contrasto ai Discorsi e ai Fenomeni d'Odio website: https://www.retecontrolodio.org/2021/10/26/ddl-zan-voto-senato-sia-palese/ [Accessed 26 January 2022]
- Sherloc United Nations Office on Drugs and Crime website: https://sherloc.unodc.org/cld/en/legislation/deu/german_criminal_code/special_part_-_chapter_seven/section_130/section_130.html [Accessed 5 February 2022]
- The Free Speech Center Website: https://www.mtsu.edu/first-amendment/article/189/brandenburg-v-ohio [Accessed 29 January 2022]
- The Future of Free Speech website: https://futurefreespeech.com/hate-speech-case-database/ [Accessed 20 January 2022]
- Twitter terms of service website: https://twitter.com/en/tos [Accessed 16 January 2022]
- Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/che-cos-e-unar [Accessed 23 January 2022]
- Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/la-mission-di-unar [Accessed 23 January 2022]
- Ufficio Nazionale Antidiscriminazioni Razziali website: https://unar.it/portale/web/guest/rete-antidiscriminazione [Accessed 23 January 2022]
- Ufficio Nazionale Antidiscriminazioni Razziali website: https://www.unar.it/portale/faq [Accessed 23 January 2022]
- UK Parliament website: https://www.parliament.uk/about/living-heritage/transformingsociety/private-lives/relationships/collections1/race-relations-act-1965/race-relations-act-1965/ [Accessed 9 January 2022]
- United Nations Free and Equal website: https://www.unfe.org/about-2/ [Accessed 17 January 2022]
- United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/HRBodies/CERD/Pages/CERDIntro.aspx [Accessed 22 November 2021].
- United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/HRBodies/CERD/Pages/EarlyWarningProcedure.aspx [Accessed 22 November 2021].
- United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/EN/Issues/SexualOrientationGender/Pages/GenderTheory.aspx [Accessed 26 January 2022]
- United Nations Human Rights Office of the High Commissioner website: https://www.ohchr.org/en/issues/freedomopinion/pages/opinionindex.aspx [Accessed 2 December 2021].
- United Nations Human Rights Office of the High Commissioner website: https://tbinternet.ohchr.org/_layouts/15/TreatyBodyExternal/Treaty.aspx?CountryID=62&Lang=EN [Accessed 18 January 2022]

- United Nations website: https://www.un.org/sg/en/content/sg/statement/2019-06-18/secretary-generals-remarks-the-launch-of-the-united-nations-strategy-and-plan-of-action-hate-speech-delivered [Accessed 15 November 2021].
- United Nations website: https://www.un.org/WCAR/e-kit/backgrounder1.htm [Accessed 15 November 2021].
- Wired website: https://www.wired.it/attualita/politica/2021/04/28/ddl-zan-calendarizzato-senato/ [Accessed 26 January 2022]
- Wired website: https://www.wired.it/attualita/politica/2021/05/04/ddl-zan-articolo-spiegato/ [Accessed 25 January 2022]