

UNIVERSITY OF GRAZ

**European Master's Programme in Human Rights and
Democratisation
A.Y. 2024/2025**

**'Slept Her Way to the Top' and a Million
Slurs**

Gendered Hate Speech and Disinformation Targeting Female Politicians
in European Legal Perspective

Author: Beatriz Castro

Supervisors: Univ.-Prof. Mag. Dr.iur. Wolfgang Benedek & Ambassador Prof. Dr.
Helmut Tichy

Word Count Declaration: 29.724

Abstract

Violence against women in politics is not a recent phenomenon, but its impact has intensified with the rise of digital technology. Among its most harmful manifestations are gendered hate speech and gendered disinformation, posing serious threats to human rights and democracy. This thesis examines how European legal frameworks address these phenomena, focusing on two main aspects: conceptualising their scope and impact, and analysing legal responses within instruments of the Council of Europe and the European Union, including jurisprudence of the European Court of Human Rights. The research highlights the difficulty of defining the boundaries of gendered hate speech, largely due to the lack of universal standards and recognition. This creates uncertainty about when gendered speech should be excluded from free speech protections, particularly in political contexts where those protections are especially robust. Gendered disinformation is identified as a distinct yet interconnected phenomenon, whose hybrid nature presents intricacies for legal classification and regulation. These challenges are echoed in the governance of social media platforms, where the EU Digital Services Act offers progress but remains constrained by vague definitions, uneven enforcement, and the evolving role of Artificial Intelligence. This thesis ultimately calls for comprehensive, gender-sensitive legal frameworks to effectively address these threats.

Key words: gendered hate speech, gendered disinformation, violence against women in politics, social media, Council of Europe, EU.

Acknowledgements

Writing this thesis was both a demanding and deeply rewarding journey – one that has challenged me intellectually and enriched me personally. It would not have been possible without the support, guidance, and inspiration of several individuals to whom I am profoundly grateful.

First and foremost, I extend my sincere gratitude to my supervisors, Professor Wolfgang Benedek and Ambassador Professor Helmut Tichy. Their generous guidance and invaluable feedback were instrumental in shaping the substance of this work.

I am also deeply thankful to the University of Graz, in particular, the European Training and Research Centre for Human Rights and Democracy, for warmly welcoming me during my second semester. Special thanks go to Mag.iur. Elias Faller and Professor Gerd Oberleitner for their generous support and for making Graz feel like home. I owe particular appreciation to Mag.iur. Gregor Fischer for kindly sharing his insights from previous projects, which played an important role in shaping the direction of this work.

My deepest gratitude goes to my family. To my parents, for making this journey possible and for their unwavering support throughout; a special thanks goes to my father, for generously taking the time to read parts of this work and provide me with invaluable advice. To my sister Clara, for giving me the privilege of sharing the E.MA experience with my best friend. And to my grandmother Beatriz, for being a constant source of strength in my life.

Lastly, I am grateful to my fellow E.MA colleagues – for their cherished friendship throughout the year, and for inspiring me with their unique sense of mission.

Table of Contents:

Chapter I: Introduction.....	6
1.1. <i>Research Question</i>	8
1.2. <i>Relevance</i>	8
1.3. <i>Methodology and Structure</i>	10
1.4. <i>Scope and Limitations</i>	11
Chapter II: Conceptualising Violence Against Women in Politics.....	12
2.1. <i>Definition</i>	12
2.2. <i>The Online Sphere</i>	17
2.2.1. <i>Gendered Hate Speech</i>	22
2.2.2. <i>Gendered Disinformation</i>	32
Chapter III: Freedom of expression v. Hate Speech: Where Does Gendered Discourse Fit Into the Debate?	39
3.1. <i>The Council of Europe</i>	41
3.1.1. <i>The Specificity of Political Contexts</i>	51
3.2. <i>The EU</i>	55
3.2.1. <i>Towards expanding the scope of gender-based cyberviolence to include sexist hate speech</i>	57
3.2.2. <i>The European Commission’s Proposal to Extend the list of ‘EU crimes’ to include Hate Speech (2021)</i>	59
3.3. <i>Gendered Disinformation: at the Crossroads of Protected and Unprotected Speech</i>	65
Chapter IV: Regulating Gendered Hate Speech and Disinformation on Social Media.....	68
4.1. <i>The EU Digital Services Act</i>	70
4.1.1. <i>Defining ‘Illegal Content’</i>	71
4.1.2. <i>Due Diligence Obligations</i>	73
4.1.3. <i>Content Removal Obligations</i>	77
4.1.4. <i>The Obligation to Remove ‘Identical’ or ‘Equivalent’ Content</i>	78
4.2. <i>The Role of Artificial Intelligence</i>	81
Chapter V: General Conclusions.....	87
Bibliography	90

List of Abbreviations

AI Artificial Intelligence

App Application

Art Article

CFREU Charter of Fundamental Rights of the European Union

CEDAW Convention on the Elimination of All Forms of Discrimination Against Women

CoE Council of Europe

DSA Digital Services Act

ECHR European Convention on Human Rights

ECJ European Court of Justice

ECtHR European Court of Human Rights

EIGE European Institute For Gender Equality

EU European Union

GR General Recommendation

GREVIO Group of Experts on Action against Violence against Women and Domestic Violence

ICT Information and Communication Technologies

MP Member of Parliament

OVAW Online Violence Against Women

Para Paragraph

TEU Treaty on European Union

TFEU Treaty on the Functioning of the European Union

UN United Nations

VAWIP Violence Against Women in Politics

VLOP Very Large Online Platform

Chapter I: Introduction

Her wings are cut and then she is blamed for not knowing how to fly.

Simone de Beauvoir

In recent years, online hate speech and disinformation targeting female politicians has emerged as a threat to both human rights and democracies worldwide.¹ Compounded by the radicalisation of public discourse and the amplifying role of technology, violence against women in politics (VAWIP) has become a stark indicator of persistent gender inequality and democratic fragility, even within liberal democracies.² European states are no exception.

While political violence is generally on the rise in Europe, its expression often shifts when the target is a woman – taking on sexualised, dehumanising, and gendered tones, aimed not at challenging political views, but at delegitimising women’s presence in politics altogether.³ A 2018 Inter-Parliamentary Union (IPU) study, based on voluntary one-to-one conversations with 123 women from 45 European countries – 81 Members of Parliament (MPs) and 42 members of parliamentary staff – provides compelling evidence

¹ Ipsos, *Survey on the Impact of Online Disinformation and Hate Speech* (UNESCO, September 2023) 1-25. See also Commissioner for Human Rights, Council of Europe, *Hate Speech Against Women Should Be Specifically Tackled* (Human Rights Comment, 6 March 2014).

² Pramukh Nanjundaswamy Vasist, Debashis Chatterjee and Satish Krishnan, ‘The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-Country Configurational Narrative’ (2023) *Information Systems Frontiers* 1, 1; Ionel Zamfir, *Violence against Women Active in Politics in the EU: A Serious Obstacle to Political Participation* (European Parliamentary Research Service, 2024) 1; Ellen Judson, ‘Gendered Disinformation: 6 Reasons Why Liberal Democracies Need to Respond to This Threat’ (Heinrich Böll Stiftung, 9 July 2021) https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat#_ftn83 accessed 12 May 2025.

³ ODIHR and OSCE, *Addressing Violence against Women in Politics in the OSCE Region Toolkit: Tool 2 – Addressing Violence against Women in Parliaments* (ODIHR and OSCE 2022) 3; Jamine Bigio and Rachel Vogelstein, ‘Women Under Attack: The Backlash Against Female Politicians’ (2020) 99 *Foreign Affairs* 131, 134; Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019)1 including Appendix, section II.B.1. See also Jana Boukemia, Marius Sältzer, Sébastien Boyer, ‘Misogyny, politics, and social media determinants of hostile engagement against women parliamentarians on Twitter’ (Legislative Studies Quarterly, 2024).

of this reality.⁴ The large majority of interviewed MPs (85.2 per cent) had faced some form of psychological violence during their term of office: 46.9 per cent had received death threats or threats of rape or beating; 58.2 per cent had been the target of online sexist attacks on social networks; and 67.9 per cent had been subjected to comments relating to their physical appearance or based on gender stereotypes.⁵

The severity and frequency of such violence are further exacerbated by intersectional discrimination – often based on race, ethnicity or ‘being young’⁶ – as Kimberlé Crenshaw insightfully observed decades ago.⁷

In this context, gendered hate speech and gendered disinformation constitute two interconnected and persistent forms of abuse that impact women in politics on a daily basis.⁸

How far should the protection of freedom of expression go when political criticism takes on a gendered, discriminatory nature? This thesis investigates how European legal frameworks address gendered hate speech and disinformation as forms of VAWIP.

⁴ Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians in Europe* (IPU 2018) 1.

⁵ Ibid.

⁶ Ibid 2, 6; Sofía Collignon, Rosie Campbell and Wolfgang Rüdiger, ‘The Gendered Harassment of Parliamentary Candidates in the UK’ (2022) 93(1) *Political Quarterly* 32, 35.

⁷ Kimberlé Crenshaw, ‘Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color’ (1991) 43 *Stanford Law Review* 1241, 1243; Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians in Europe* (IPU 2018) 2, 6. See also: Amnesty International, ‘Diane Abbott: Violence Against Women Online’ (21 March 2018) <https://www.amnesty.org/en/latest/news/2018/03/diane-abbott-online-violence-against-women/> accessed 1 May 2025.

⁸ See, for example: Kristina Van Sant, Rolf Fredheim and Gundars Bergmanis-Korāts, *Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers* (NATO StratCom Centre of Excellence, February 2021); Amnesty International, ‘Diane Abbott: Violence Against Women Online’ (21 March 2018) <https://www.amnesty.org/en/latest/news/2018/03/diane-abbott-online-violence-against-women/> accessed 1 May 2025; Lucina Di Meco, ‘Campaigning while female: How online hate holds women back’ (Politico, 24 September 2021) <https://www.politico.eu/article/campaigning-while-female-how-online-hate-holds-women-back/> accessed 1 May 2025; Laura Boldrini, ‘Digital Misogyny and Character Assassination’ (*Foundation for European Progressive Studies*, 24 November 2020) <https://feps-europe.eu/digital-misogyny-and-character-assassination/> accessed 1 May 2025; Amnesty International, ‘Nicola Sturgeon: A Case Study in Violence Against Women Online’ (21 March 2018) <https://www.amnesty.org/en/latest/news/2018/03/nicola-sturgeon-online-violence-against-women/> accessed 1 May 2025; Raquel Martínez-Sanz, Patricia Durántez-Stolle and Iris Simón-Astudillo, ‘Memes as Hate Speech: Violence, Humour and Criticism Surrounding the Image of Irene Montero’ (2024) 16(1) *International Visual Culture Review* (Universidad de Valladolid, España); Goldberg E, ‘Fake Nudes and Real Threats: How Online Abuse Holds Back Women in Politics’ *The New York Times* (4 June 2021) <https://www.nytimes.com/2021/06/03/us/disinformation-online-attacks-female-politicians.html> accessed 6 May 2025.

1.1. Research Question

To what extent do European legal frameworks address gendered hate speech and gendered disinformation targeting women in politics, particularly on social media platforms?

1.2. Relevance

Although a growing number of women are breaking the ‘glass ceiling’ in politics and stepping into leadership roles, their overall representation continues to fall short from parity with men. Nearly three decades after the Beijing Declaration and Platform for Action advocated for increased female political participation and the removal of barriers to such engagement,⁹ progress remains limited and fragile. A recent IPU report shows that women's representation in national parliaments rose from 11 per cent in 1995 to 27.2 per cent in 2015.¹⁰ Although there was steady progress between 2000 and 2015, the pace of growth has slowed in more recent years.¹¹ The radicalisation of public discourse online has visibly fuelled a backlash against women in such decision-making positions.¹² In this context, VAWIP remains a major deterrent, particularly due to its chilling effect.¹³

A news report reveals that 1 in every 6 women worldwide leaves politics during their first term serving.¹⁴ Many cite misogynistic abuse as a major factor in their decision to step

⁹ United Nations, *Beijing Declaration and Platform for Action* (UN, 1995) paras 1, 10, 181-189.

¹⁰ Inter-Parliamentary Union, *Women in parliament: 1995-2025* (IPU 2025) 2.

¹¹ Ibid.

¹² Rosa Celorio, *Women and International Human Rights in Modern Times: A Contemporary Casebook* (Edward Elgar Publishing, 2022) 176; Christine Ro, ‘Why Do We Still Distrust Women Leaders?’ *BBC News* (19 January 2021) <https://www.bbc.com/worklife/article/20210108-why-do-we-still-distrust-women-leaders> accessed 3 June 2025; UN Women, ‘Facts and Figures: Women’s Leadership and Political Participation’ (UN Women, 2025) <https://www.unwomen.org/en/articles/facts-and-figures/facts-and-figures-womens-leadership-and-political-participation> accessed 21 March 2025.

¹³ Ionel Zamfir, *Violence against Women Active in Politics in the EU: A Serious Obstacle to Political Participation* (European Parliamentary Research Service, 2024) 1.

¹⁴ Věra Jourová, ‘Europe Must Do More to Keep Women in Public Life’ *Financial Times* (27 May 2024) <https://www.ft.com/content/418dfc3e-f113-436f-9cc3-e28df1b76ca3> accessed 30 June 2025.

down.¹⁵ Alarming, this hostile online environment risks escalating into offline violence, including physical attacks. A stark example was the brutal murder of British Member of Parliament (MP) Jo Cox by a white supremacist in 2016.¹⁶

In the early stages of my research, I found that scholarly literature increasingly recognises the growing threats that online gendered attacks pose to democratic values and gender equality globally. However, limited academic texts specifically addressed VAWIP in the digital sphere as a distinct and pressing issue – a gap that becomes even more pronounced when looking solely at the European context.¹⁷

On the one hand, most research concentrated on how this violence undermines electoral integrity during election periods, with the majority of VAWIP accounts coming from countries with weaker democratic institutions.¹⁸ This limited approach, however, does not fully address the broader implications of gendered political violence. Research has demonstrated that this phenomenon is not confined to authoritarian or developing states, and its impact is not limited to electoral periods, but rather, it is deeply entrenched within the political and societal fabric.¹⁹

On the other hand, while online gender-based violence, including hate speech, is a long-standing topic of discussion, it is still underreported in Europe – and even more so when looking at the specific case of women in politics –, making systematic analysis of the issue particularly challenging.²⁰

¹⁵Global Disinformation Index, *Gendered Disinformation in the European Parliamentary Elections* (10 June 2024) <https://www.disinformationindex.org/blog/2024-06-10-gendered-disinformation-in-the-european-parliamentary-elections/> accessed 6 March 2025; Hope not Hate, 'The Freedom to Be Free from Abuse' (no date) <https://hopenothate.org.uk/in-search-of-hope/the-freedom-to-be-free-from-abuse/> accessed 16 June 2025.

¹⁶ BBC News, 'Labour MP Jo Cox murdered for political cause' (BBC News, 14 November 2016) <https://www.bbc.com/news/uk-37978582> accessed 25 April 2025 accessed 6 May 2025; Mona Lena Krook and Juliana Restrepo Sanín, 'The Cost of Doing Politics? Analyzing Violence and Harassment against Female Politicians' (2020) *Perspectives on Politics* 18(3) 740, 751-753.

¹⁷ Malin Holm, 'Violence against Women in Politics: Emerging Perspectives, New Challenges' (2020) 3(2) *European Journal of Politics and Gender* 295, 297.

¹⁸ Sandra Håkansson, 'Do Women Pay a Higher Price for Power? Gender Bias in Political Violence in Sweden' (2021) 83(2) *The Journal of Politics* 515, 517, 528.

¹⁹ Mona Lena Krook and Juliana Restrepo Sanín, 'Violence Against Women in Politics. A Defence of the Concept' (2016) *Política y Gobierno* 23(2), 3.

²⁰ Kim Barker and Olga Jurasz, 'Online Violence Against Women as a Key Obstacle to Gender Equality: A Critical View from Europe' (2020) 1 *European Equality Law Review* 47, 4.

What is more, although disinformation disseminated on social media has become a major threat for female politicians,²¹ a scoping review analysing studies on disinformation from 2013 to 2023 found that only 14 out of 143 articles directly addressed the gender dimensions of the phenomenon, highlighting a significant research gap.²²

From this initial research it was possible to conclude that legal approaches and policies are still primarily focusing on online gender-based violence as an issue of harassment and on disinformation as a possible threat to public order or democracy.²³ As a result, they often fail to comprehensively address the distinct legal challenges posed by hate speech and disinformation targeting women based on gender – and particularly in the specific context of politics.²⁴

Lastly, as technology is advancing faster than both legal research and regulation, allowing artificial intelligence (AI)-facilitated spread of hate speech and disinformation, significant gaps emerge in understanding how to effectively address these threats while protecting democratic participation and freedom of expression in the digital sphere.²⁵ Thus, analysing social media regulation in light of present and emerging challenges is both pertinent and necessary.

1.3. Methodology and Structure

²¹Lucina Di Meco and Saskia Brechenmacher, 'Tackling Online Abuse and Disinformation Targeting Women in Politics' (*Carnegie Endowment for International Peace*, 30 November 2020), <https://carnegieendowment.org/research/2020/11/tackling-online-abuse-and-disinformation-targeting-women-in-politics?lang=en> accessed 3 June 2025; Kristina Wilfore, 'Security, Misogyny, and Disinformation Undermining Women's Leadership' in Gulizar Hacıyakupoglu and Yasmine Wong (ed), *Gender and Security in the Digital Space* (Routledge 2022) 124, 125; Gabrielle Bardall, 'Nasty, Fake and Online: Distinguishing Gendered Disinformation and Violence Against Women in Politics' in Gulizar Hacıyakupoglu and Yasmine Wong (eds), *Gender and Security in Digital Space: Navigating Access, Harassment, and Disinformation* (1st edn, Routledge 2022) 109, 110-111.

²²Juliana Alcantara and Juliana Valentim, 'Gender-based Disinformation: A Scoping Review of the Literature, 2013-2023' (2023) 48 *ex æquo* 125, 125.

²³ Ibid 134; Taylor Agajanian and Rachel E Moran, 'Gender as a Central Site of Inquiry within Mis- and Disinformation Studies' (2024) 3 *Feminist Media Studies* 1, 1.

²⁴ Ibid.

²⁵ Noémi Bontridder and Yves Pouillet, 'The Role of Artificial Intelligence in Disinformation' (2021) 3: e32 *Data & Policy* 1, 1.

This thesis employs a socio-legal methodology grounded in literature-based research, drawing on both primary and secondary sources. These include academic books and journal articles, human rights treaties (United Nations and Council of Europe), EU legislation, soft law instruments such as reports and general recommendations, jurisprudence – particularly from the European Court of Human Rights – and media sources, including official websites and news portals.

The thesis is structured in five chapters:

- (i) Chapter I introduces and contextualises the research topic.
- (ii) Chapter II defines and conceptualises VAWIP, with emphasis on its online manifestations, particularly gendered hate speech and gendered disinformation.
- (iii) Chapter III situates gendered hate speech and gendered disinformation within the broader legal tension between protecting freedom of expression and regulating hate speech by analysing the relevant legal frameworks and jurisprudence of the ECtHR, with a particular focus on their application in cases involving female politicians.
- (iv) Chapter IV examines how social media regulation – particularly through the EU Digital Services Act – addresses gendered hate speech and disinformation, and additionally reflects on the double-edged role of AI in both amplifying and combating these phenomena.
- (v) Chapter V concludes the thesis by reflecting on the main findings.

1.4. Scope and Limitations

Primarily, it is crucial to clarify what I understand as ‘women in politics’ for the purpose of this research. Adopting a broad definition, I will address ‘women in politics’ as women that are actively engaged in political activities whether at the national, local, or regional level, independently or through a political party. This includes both high profile politicians (e.g. Prime-Ministers or Party Leaders) and low-profile (e.g. Members of

Parliament) and excludes political activists that are not formally involved in governance or electoral politics, since the focus of this research is on individuals holding or seeking political office.

Secondly, while recognising ongoing debates around the legal definition of 'woman' and the conceptual distinctions between 'sex' as a biological attribute and 'gender' as a social construct – as well as the importance of an inclusive, non-binary perspective – this thesis adopts a binary framework, using the terms 'sex' and 'gender' interchangeably for analytical purposes. Accordingly, the adjectives 'gendered', 'gender-based' and 'sexist' are used without distinction. Furthermore, despite acknowledging the intersectional aspects of VAWIP, their specific dynamics are not the primary focus of this research.

Thirdly, VAWIP encompasses various forms of gender-based violence; however, I will focus the legal analysis on the hate speech and disinformation phenomena for the reasons explained in section 1.2.

Finally, this thesis is limited to the European legal perspective, focusing of the analysis of the relevant legal frameworks the Council of Europe (CoE), and the European Union (EU). However, UN legal frameworks and mechanisms will also be examined, where relevant, to provide contextual grounding.

Chapter II: Conceptualising Violence Against Women in Politics

2.1. Definition

No international or European human rights treaty explicitly defines VAWIP. Nevertheless, interpretative efforts and recommendations found in non-binding instruments, research reports, and academic literature have sought to develop a comprehensive conceptualisation of the phenomenon.

The concept first emerged through the writing of Mona Lena Krook and Juliana Restrepo Sanín in the mid-2010s.²⁶ In her book *Violence Against Women in Politics* (2020), Krook later identifies VAWIP as a ‘distinct phenomenon’ that transcends the general risks inherent to political life, often referred to as ‘the cost of doing politics’.²⁷ This is unlike authors such as Jennifer M. Piscopo, who argue that VAWIP is more closely tied to imperfect democratic consolidation processes than to backlash against female politicians – and therefore view VAWIP as a subset of political violence.²⁸ Conversely, Krook emphasises that VAWIP needs to be distinguished from general political violence in that it is not primarily about political confrontation as such, but rather the systematic discrimination of women in political spaces on the basis of their gender.²⁹

Accordingly, an analysis of the phenomenon that relies solely on a general political violence approach – without accounting for its gendered dimension – inevitably fails to capture its true nature and underlying dynamics.

Such dynamics are reflected in three interconnected levels at which this violence operates: structural, cultural, and symbolic.³⁰ Structurally, relating to its origins, gender inequality is embedded in institutions and societal norms, limiting women's access to political participation. Culturally, relating to the means through which it operates, gender-motivated political violence is often seen as justified or ignored due to entrenched patriarchal norms (e.g. the normalization of commenting on a woman’s physical appearance). Symbolically, relating to the outcomes of such violence, the idea of ‘male primacy’ legitimises gendered violence as a means of maintaining the status quo.³¹

²⁶ See Krook and Restrepo Sanín (n 19).

²⁷ Mona Lena Krook, *Violence Against Women in Politics* (Oxford University Press 2020) 75.

²⁸ See Jennifer M Piscopo, ‘States as Gender Equality Activists: The Evolution of Quota Laws in Latin America’ (2015) 57(3) *Latin American Politics and Society* 27-49; Krook and Restrepo Sanín (n 19) 2.

²⁹ Krook and Restrepo Sanín (n 19) 3.

³⁰ Krook (n 27) 97-103.

³¹ Ibid. Pierre Bourdieu understands symbolic violence as violence that relies on the perception of legitimacy and the subsequent consent and complicity of actors concerned (Pierre Bourdieu, ‘Symbolic Power’, 1979, 4 *Critique of Anthropology* 77, 77).

The concept of VAWIP then gained prominence as organizations like the United Nations (UN),³² the IPU,³³ or the National Democratic Institute³⁴ gradually began advocating for its recognition as a distinct form of violence, investing in specialised studies and reports.

VAWIP as gender-based violence

Having distinguished VAWIP from general political violence, it becomes crucial to frame it as gender-based violence. It is here pertinent to examine early developments at the UN level, as these have subsequently informed European policy and legal frameworks.

In 1993, the UN General Assembly defined violence against women as ‘any act of gender-based violence that results in, or is likely to result in physical, sexual, or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life’.³⁵ This definition would later be reproduced in Article 3(a) of the 2011 CoE Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention).³⁶

Despite multilateral efforts to establish the elimination of all forms of gender-based violence as a priority,³⁷ the absence of a UN treaty exclusively focusing on gender-based violence still hinders further attempts to effectively address the issue at policy level. Although the UN Convention on the Elimination of All Forms of Discrimination Against

³² Particularly through UN Women, the Committee on the Elimination of Discrimination Against Women, the Special Rapporteur on Violence Against Women and the UNGA Res 73/148 *Intensification of efforts to prevent and eliminate all forms of violence against women and girls: sexual harassment* (17 December 2018) UN Doc A/RES/73/148, which in para 7 specifically addresses violence against women in politics, calling on states to take action.

³³ See Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians* (IPU 2016).

³⁴ Specially through the #NotTheCost campaign (2016) and its renewed call to action in 2021, which explicitly framed VAWIP as a global issue of human rights and democracy (National Democratic Institute, *#NotTheCost: Stopping Violence Against Women in Politics: A Call to Action*, NDI 2016; National Democratic Institute, *#NotTheCost: Stopping Violence Against Women in Politics: A Renewed Call to Action*, NDI 2021).

³⁵ UNGA Res 48/104, *Declaration on the Elimination of Violence Against Women* (20 December 1993) UN Doc A/RES/48/104, Art 1.

³⁶ Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention) (adopted 11 May 2011, entered into force 1 August 2014) CETS No 210, Art 3(a).

³⁷ *Vienna Declaration and Programme of Action*, World Conference on Human Rights, Vienna, 14-25 June 1993, UN Doc A/CONF.157/24 (Part I) (1993) Art 18; United Nations, *Beijing Declaration and Platform for Action* (UN, 1995); Rosa Celorio, *Women and International Human Rights in Modern Times: A Contemporary Casebook* (Edward Elgar Publishing, 2022) 38.

Women (CEDAW) has promoted important progress in advancing the prohibition of discrimination based on gender, its text does not cover the issue of gender-based violence and only in 1992 did its monitoring body establish an intricate link to ‘discrimination against women’.³⁸

Indeed, to fill the treaty’s gap in this regard, the Committee issued three General Recommendations (GR) – No. 12, 19, and 35 (the latter as an update to GR no. 19) – that address gender-based violence and clarify its link to the prohibition of discrimination under CEDAW.³⁹

Notably, GR No. 19 (1992) served as a pivotal force for the evolution of the prohibition of gender-based violence into a principle of customary international law.⁴⁰ In its authoritative interpretation of Article 1 CEDAW, the Committee defined gender-based violence against women for the first time as ‘violence that is directed against a woman because she is a woman or that affects women disproportionately’.⁴¹ This definition was later echoed at the European level in Article 3(d) of the Istanbul Convention and, more recently, in Directive (EU) 2024/1385 on Combating Violence Against Women and Domestic Violence, in Article 2(1).⁴² Through its dual structure, this formulation recognises that violence against women may not always stem from explicit gender bias, yet can still be considered ‘gendered’ insofar as it disproportionately affects women and reinforces systemic gender inequality. This nuanced understanding is crucial for identifying VAWIP as gender-based violence, since it can sometimes be disguised as legitimate political criticism.

³⁸ Committee on the Elimination of Discrimination against Women, *General Recommendation No 19: Violence against women* (30 January 1992) UN Doc A/47/38, para 1.

³⁹ Rosa Celorio, *Women and International Human Rights in Modern Times: A Contemporary Casebook* (Edward Elgar Publishing, 2022) 39.

⁴⁰ Committee on the Elimination of Discrimination against Women, *General Recommendation No 35 on Gender-Based Violence against Women, Updating General Recommendation No 19* (14 July 2017) UN Doc CEDAW/C/GC/35, para 2.

⁴¹ Committee on the Elimination of Discrimination against Women, *General Recommendation No 19: Violence against women* (30 January 1992) UN Doc A/47/38, para 6.

⁴² Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention) (adopted 11 May 2011, entered into force 1 August 2014) CETS No 210, Art 3(d); Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Art 2(1).

While GR No. 19 (1992) already affirmed that underlying consequences of gender-based violence helped maintaining a ‘low level of political participation’, its updated version in GR No. 35 (2017) explicitly recognised harmful practices and crimes against politicians as a form of gender-based violence against women, acknowledging that the phenomenon is affected by ‘cultural, ideological and political factors’.⁴³ Importantly, the latter GR further underscored that gender-based violence is not confined to a specific setting: it can occur anywhere – public or private – including in online spaces and within political contexts.⁴⁴

Additionally, the root causes of gender-based violence against women warrant particular attention from the Committee, as reflected in its references to the need to challenge ‘patriarchal attitudes’ and ‘gender stereotypes’.⁴⁵ In fact, the very text of CEDAW already mandated states to modify the social and cultural patterns of conduct of men and women, and to eliminate prejudices and practices rooted in stereotyped notions of gender roles, including within the educational context.⁴⁶

A working definition

In 2016 the IPU issued a crucial report on ‘Sexism, harassment and violence against women parliamentarians’. The study was based on quantitative and qualitative data provided voluntarily by 55 women parliamentarians from 39 countries all over the world, collected on the occasion of the international campaign *#NotTheCost, Stopping Violence against Women in Politics*. Although the report was published nearly a decade ago, it remains highly relevant, particularly in its identification of three key characteristics of VAWIP:

- (i) It targets women because of their gender;
- (ii) In its very form it can be gendered, as exemplified by sexist threats and sexual violence;

⁴³ Committee on the Elimination of Discrimination against Women, ‘General Recommendation No 35 on Gender-Based Violence against Women, Updating General Recommendation No 19’ (14 July 2017) UN Doc CEDAW/C/GC/35, para 14.

⁴⁴ *Ibid* para 20.

⁴⁵ *Ibid* para 34.

⁴⁶ Convention on the Elimination of All Forms of Discrimination against Women (adopted 18 December 1979, entered into force 3 September 1981) 1249 UNTS 13, Arts 5(a), 10(c).

(iii) Its impact is to discourage women from being or becoming active in politics.⁴⁷

However, point (iii) introduces an effects-based criterion that arguably sets the threshold for identifying VAWIP excessively high. By requiring evidence of discouragement of women's political participation, it risks excluding forms of violence or harassment that, while potentially harmful, may not result in immediately measurable political withdrawal. I therefore propose reformulating point (iii) to require that the violence be 'capable of' discouraging women from entering or remaining in politics, rather than demanding proof of its actual effect.

Building on what has been said, and while a consensual definition of VAWIP remains contested and underdeveloped, I propose the following working definition: VAWIP encompasses any act of abuse, whether physical or psychological, verbal or digital (including text-based, visual, or audiovisual content), that targets women in political life primarily or significantly because of their gender. Such acts may take explicitly gendered forms, such as sexist insults and threats or sexual violence, or may appear gender-neutral but disproportionately affect women compared to men in the same position. Crucially, these are capable not only of causing individual harm but also undermining women's political participation and freedom of expression more broadly, thereby reinforcing structural gender inequality.

2.2. The Online Sphere

The radicalisation of political discourse and the 'online disinhibition effect' in social media – enabled by anonymity and a perception of impunity – is contributing to the spread of misogynistic expression.⁴⁸

⁴⁷ Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians* (IPU 2016) 2; National Democratic Institute, *#NotTheCost: Stopping Violence Against Women in Politics: A Renewed Call to Action* (NDI 2021) 12.

⁴⁸ Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Recital 25; Gabrielle Bardall, 'Nasty, Fake and Online: Distinguishing Gendered Disinformation and Violence Against Women in Politics' in

Online violence against women (OVAW) often referred to interchangeably as ‘gender-based cyberviolence’ or ‘technology facilitated gender-based violence’⁴⁹ has been formally recognised as a major global threat.⁵⁰ In 2018, the UN Special Rapporteur on violence against women was already drawing attention to the role of social media platforms and other digital spaces as vectors for gender-motivated attacks on women in public life, as set out in a seminal report on VAWIP.⁵¹ At the European level, the CoE’s Gender Equality Strategy (2024-2029) underscores the importance of tackling both the root causes and digital manifestations of VAWIP and commits to combating sexist hate speech, violent and sexualised threats, doxing and trolling on social media platforms.⁵² CoE’s Recommendation on Combating Hate Speech further calls on Member States to implement reporting mechanisms for online hate speech,⁵³ while its Recommendation on Preventing and Combating Sexism addresses structural stereotypes that fuel gendered abuse online.⁵⁴ The EU Gender Equality Strategy (2020-2025) further warns that online abuse of women has reached an ‘unacceptable’ level with ‘specific’ and ‘vicious’ effects on women’s public participation.⁵⁵

However, fragmented national frameworks, inconsistent definitions – or lack thereof –, and the constant evolution of new forms of abuse undermine legal and policy coherence at the European level, resulting in significant protection and enforcement gaps in the field of OVAW⁵⁶ – gaps that become even more pronounced when focusing on the specific case of women in politics.

Gulizar Hacıyakupoglu and Yasmine Wong (eds), *Gender and Security in Digital Space: Navigating Access, Harassment, and Disinformation* (1st edn, Routledge 2022) 109, 115.

⁴⁹ For instance, GREVIO uses the term ‘digital dimension of violence against women’ to refer to both online acts of violence and those perpetrated through technology, including technology yet to be developed (European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls*, EIGE 2022, 66).

⁵⁰ European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls* (EIGE 2022) 66-84.

⁵¹ Dubravka Šimonović, *Report of the Special Rapporteur on Violence against Women in Politics* (6 August 2018) UN Doc A/73/301, para 83(a).

⁵² Council of Europe, *Gender Equality Strategy (2024-2029)*, paras 47-54.

⁵³ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, paras 19-27.

⁵⁴ Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019)1 including Appendix.

⁵⁵ European Commission, ‘A Union of Equality: Gender Equality Strategy 2020-2025’ COM(2020) 152 final, 5 March 2020, 5.

⁵⁶ *Ibid*; Barker and Jurasz (n 20) 2.

Notably, at present, EU Member States typically address cyberviolence through one of four legal frameworks: (i) subsuming it under general offences without any reference to Information and Communication Technologies (ICT); (ii) incorporating ICT-related elements into existing general offences; (iii) treating cyberviolence as an aggravating factor in broader criminal statutes; or (iv) enacting dedicated cyberviolence provisions. Yet only a handful of Member States have adopted standalone cyberviolence laws, and even these instruments are largely gender-neutral, omitting any explicit reference to violence against women or other gender-based harms.⁵⁷

Regarding the definitional issue, it is worthy of note that neither CEDAW, nor the Istanbul Convention, two leading international treaties on human rights of women, explicitly reference online forms of gender-based violence in their respective texts.⁵⁸

Nevertheless, the respective monitoring bodies – the CEDAW Committee and the Group of Experts on Action Against Violence Against Women and Domestic Violence (GREVIO) – have emphasised the pressing necessity to address the digital dimension of violence against women and have advanced with efforts to define the phenomenon, encompassing the context of politics.⁵⁹

At the UN level, the CEDAW Committee clarified, in GR No. 35 (2017), that gender-based violence includes any ICT-facilitated act causing physical, sexual or psychological harm to women.⁶⁰ In 2018 the UN Special Rapporteur on Violence Against Women clearly defined gender-based cyberviolence as:

‘any act of gender-based violence against women that is committed, assisted or aggravated in part or fully by the use of ICT, such as mobile phones and smartphones, the Internet, social media

⁵⁷ European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls* (EIGE 2022) 24-32.

⁵⁸ Notably, CEDAW does not explicitly mention gender-based violence at all, as previously discussed.

⁵⁹ CEDAW Committee, ‘General Recommendation No 35 on Gender-Based Violence against Women, Updating General Recommendation No 19’ (14 July 2017) UN Doc CEDAW/C/GC/35, para 20; Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women*, Recommendation No 1, adopted 20 October 2021, para 50).

⁶⁰ Committee on the Elimination of Discrimination against Women, ‘General Recommendation No 35 on Gender-Based Violence against Women, Updating General Recommendation No 19’ (14 July 2017) UN Doc CEDAW/C/GC/35, para 20.

platforms or email, against a woman because she is a woman, or affects women disproportionately'.⁶¹

When centring around social media platforms, OVAW includes, but is not limited to:

- (i) Text-based abuse – such as trolling,⁶² doxing,⁶³ cyberstalking, cyberbullying, cyber-harassment and offensive comments.
- (ii) Video and Image-based abuse – mostly (but not exclusively) of sexual nature, such as non-consensual sharing of intimate images or videos, including deepfakes⁶⁴ and revenge porn.⁶⁵

All these forms can arguably be vehicles of hate speech⁶⁶ or disinformation, and they disproportionately affect women in public life – such as politicians, journalists, and women's rights defenders.⁶⁷

⁶¹ Dubravka Šimonović, *Report of the Special Rapporteur on online violence against women and girls from a human rights perspective* (18 June 2018) UN Doc A/HRC/38/47, para 23.

⁶² 'Trolling' is broadly understood as defamatory and antagonistic messages targeting users of social media, with the intent to provoke an argument or simply an emotional reaction. (University of Warwick and TNO, *Workshop 5: Policing of Trolling on Social Media, Ethical and Legal Issues, Deliverable 4.5*, 2018, 4). Karla Mantilla further identifies 'gendertrolling' as a specific type of trolling that is in its very nature specifically directed at women; it is characterised by the use of misogynistic slurs, aggressive and hostile language, credible threats of violence, coordinated involvement of multiple individuals, and an exceptional degree of intensity, breadth, and persistence. Such attacks especially target women that publicly expose themselves in domains historically dominated by men (Karla Mantilla, 'Gendertrolling: Misogyny Adapts to New Media', 2013, 39 *Feminist Studies* 563, 564-565, 569).

⁶³ The term is a reference to 'dropping documents', regarding online researching and publishing of private information on the internet to publicly expose and shame the person targeted (European Institute for Gender Equality, *Gender Pay Gap*, 2024, https://eige.europa.eu/publications-resources/thesaurus/terms/1460?language_content_entity=en accessed 22 March 2025; Sarah Sobieraj, *Credible Threat: Attacks Against Women Online and the Future of Democracy*, Oxford University Press 2020, 14).

⁶⁴ 'Deepfake' is defined as 'AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful' in Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L168/1, Art 3(60).

⁶⁵ UN Women, 'Digital abuse: Trolling, stalking and other forms of technology-facilitated violence against women' (*UN Women*, 10 February 2025) <https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women> accessed 24 March 2025.

⁶⁶ See, for example, in relation to cyberbullying: Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021) para 45.

⁶⁷ *Ibid.*

EU legislative and regulatory attempts have progressively developed protections against OVAW. For instance, Directive (EU) 2024/1385 mandates States to criminalise certain forms, specifically: non-consensual sharing of intimate or manipulated material (Article 5), cyberstalking (Article 6), cyber-harassment (Article 7) and cyber-incitement to violence or hatred on grounds of gender (Article 8).⁶⁸

Still, the absence of a universal definition to date has resulted in fragmented and inconsistent approaches, featuring overlapping and imprecise terminology and thus failing the full spectrum of the phenomenon.⁶⁹

Nevertheless, GREVIO recognises that ‘not all acts of violence against women in the digital sphere are of the same severity, nor do they all meet the threshold for criminal prosecution within individual states’.⁷⁰ Indeed, following political scientist Gabrielle Bardall, it is possible to categorise forms of OVAW according to levels of intensity and harm ranging from general acts of incivility to outright threats.⁷¹ In this context, the author considers ‘online hate speech’ to be one of the most egregious forms of attack, alongside ‘direct and explicit threats of physical or sexual harm’. By contrast, disinformation is regarded as a less severe form of incivility.⁷²

Before turning to a detailed analysis of gendered hate speech and disinformation, two preliminary observations about the European legal approach to OVAW forms are necessary.

First, there remains a persistent lack of recognition of text-based abuse as a form of OVAW. A common critique is the disproportionate focus on the sexualised aspects of abuse, particularly when it involves image-based content.⁷³ This narrow emphasis tends to constrain legal and policy responses, often overlooking gender-based hate speech and

⁶⁸ Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence [2024] OJ L 1385/1, Arts 5-8.

⁶⁹ Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021) para 28.

⁷⁰ *Ibid* para 29.

⁷¹ Gabrielle Bardall, ‘Nasty, Fake and Online: Distinguishing Gendered Disinformation and Violence Against Women in Politics’ in Gulizar Hacıyakupoglu and Yasmine Wong (eds), *Gender and Security in Digital Space: Navigating Access, Harassment, and Disinformation* (1st edn, Routledge 2022) 109, 110-111.

⁷² *Ibid*.

⁷³ Barker and Jurasz (n 20) 3.

targeted disinformation, which typically appear in text-based formats – such as posts, comments, or messages on social media. As a result, the development of a comprehensive framework capable of addressing the full spectrum of gendered online abuse remains limited.⁷⁴

Second, significant challenges emerge in addressing these phenomena through platform content moderation obligations.⁷⁵ Chief among them is the difficulty of distinguishing between illegal content and harmful – but not necessarily illegal – content, while balancing the protection of freedom of expression with the obligation to address material that may harm users or society at large.⁷⁶

2.2.1. Gendered Hate Speech

‘Sexist hate speech’, as named in several CoE documents, including the Gender Equality Strategy (2024-2029), has emerged as a significant and distinct form of OVAW, endangering women involved in the public scene.⁷⁷ GREVIO’s Recommendation on the Digital Dimension of Violence Against Women highlights the severe impact of hate speech directed at women in male-dominated spheres and those who challenge traditional gender norms, underscoring its role in reinforcing structural inequality.⁷⁸ The document cites politicians, journalists, human rights defenders and activists as particularly vulnerable groups.⁷⁹

This was already stated in the CoE’s Recommendation on Preventing and Combating Sexism⁸⁰ where the existing gap in policies and legislation was acknowledged, along

⁷⁴ Ibid.

⁷⁵ Ibid 2.

⁷⁶ Ibid.

⁷⁷ See Council of Europe, *Gender Equality Strategy (2024-2029)*.

⁷⁸ Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021) paras 44, 45.

⁷⁹ Ibid.

⁸⁰ Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019) including Appendix, section II.B.

with its impact on hindering the consistent identification of misogynistic speech online as a violation of international and European human rights standards.⁸¹

Before conducting a legal analysis, it is important to gain an understanding of the issue by building a conceptual and practical basis.

Conceptualising ‘gendered hate speech’

Abusive language targeting all types of protected characteristics on social media can take various forms, such as being explicit or implicit, and being directed at specific individuals or generalised towards groups. This can result in comments that are (i) explicit and directed, (ii) explicit and generalised, (iii) implicit and directed, or (iv) implicit and generalised. This can be exemplified as follows:

<i>Subtype</i>	<i>Example</i>	<i>Citation</i>
Explicit and Directed	‘Youre one of the ugliest b*tches Ive ever fucking seen’	Kontostathis et al., 2013
Explicit and Generalised	‘So an 11 year old n*gger girl killed herself over my tweets? ^_^ that’s another n*gger off the streets!!’	Kwok and Wang, 2013
Implicit and Directed	‘Hey Brendan, you look gorgeous today. What beauty salon did you visit?’	Dinakar et al., 2012
Implicit and Generalised	‘most of them come north and are good at just mowing lawns’	Dinakar et al., 2011

Table 1 – Retrieved from Kristina Van Sant, Rolf Fredheim and Gundars Bergmanis-Korāts, *Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers* (NATO StratCom Centre of Excellence, February 2021) 8, citing Zeerak Waseem and others, ‘Understanding Abuse: A Typology of Abusive Language Detection Subtasks’ (2017) *Proceedings of the First Workshop on Abusive Language Online* 78, 80.⁸²

⁸¹ Ibid.

⁸² Even though examples (ii) and (iv) lack a clear gendered component, they still serve the purpose of illustrating the respective form of hate speech.

In this context, the social media treatment of Finland's female-majority government (2019-2023), led by Prime Minister Sanna Marin, offers valuable insight into the gender dimension of this type of abuse. According to a research report, which monitored comments posted on X (formerly Twitter) from March to July 2020, during the COVID-19 pandemic, a wave of abusive messages targeted Marin and her female ministers, questioning their competence to govern based on their age and gender. Derogatory terms such as 'lipstick girls' and 'tampax team' were used to undermine their credibility as political actors.

Across the monitoring period, the study found that female Finnish ministers were subjected to a disproportionately high volume of abusive messages, with a significant proportion of this abuse featuring both implicit and explicit sexist language. Some examples, translated from Finnish, are included in the following table:

<i>Date</i>	<i>Post</i>	<i>Target</i>	<i>Comment</i>
15 March 2020		Prime Minister Sanna Marin	@MarinSanna There's nothing more dangerous for the security threat facing Finland than the members of the government being socialist-green feminist girls with no life experience, no backbone and no understanding of security. @valtioneuvosto @STM_Uutiset @THLorg.
29 March 2020		Prime Minister Sanna Marin	F*ck off to hell you Social Democrat piece of sh*t, you're not qualified to run the country, and take the rest of the lipstick b*tches with you too.
21 April 2020		Prime Minister Sanna Marin	Sanna, you can shove that vaccine up your sticky c*nt! We're not taking it.
21 April 2020		Prime Minister Sanna Marin	@MarinSanna How do you manage to spew out such dreadful sh*t all the time, you left-wing hag? Go and bake something or load the washing machine. Empty-headed left winger. You clearly have no clue how to manage this situation.
25 June 2020		Minister of the Interior Maria Ohisalo	@MariaOhisalo Go f*ck yourself and don't threaten my family members, you liberal idiots, you f*cking whore

Table 2 — Retrieved from Kristina Van Sant, Rolf Fredheim and Gundars Bergmanis-Korāts, *Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers* (NATO StratCom Centre of Excellence, February 2021) 38-51.

These messages illustrate that regardless of the political topic at hand, gendered language was consistently used to criticise the performance of female ministers in their roles as government officials.⁸³

Nevertheless, the term 'abusive' when categorising messages on social media is ambiguous and may be considered overly broad as a means of qualifying hate speech. It is therefore imperative to be able to clearly identify where hate speech is present and differentiate it from broader online abuse. In this context, the fundamental question of when gendered expression conceptually amounts to the threshold of hate speech must be addressed prior to proceeding to the analysis of legal frameworks.

When does 'Sexist Speech' Become 'Hate Speech'?

According to social philosopher Louise Richardson-Self, how to properly situate patriarchy-enforcing speech – that is, sexist speech –, within the category of hate speech is an important politico-philosophical project.⁸⁴

In order to avoid excessive curtailment of free speech – which is particularly vital in political contexts – it is important to distinguish between sexist 'hate' speech, often referred to as misogynistic speech, and less severe forms of sexist expression, often referred to as sexist speech.

While both forms target women negatively on the basis of their gender – and are recognised as human rights issues, particularly under specialised instruments such as CEDAW and the Istanbul Convention – they impact women in distinct ways. Succinctly: sexist rhetoric can be defined as language that reflects a hierarchical social structure in support of a patriarchal gender order, with the purpose of perpetrating rooted stereotypes about gender roles – particularly the belief that men are naturally superior to women.⁸⁵ In

⁸³ Ibid 51.

⁸⁴ Louise Richardson-Self, 'Woman-Hating: On Misogyny, Sexism, and Hate Speech' (Cambridge University Press, 2017) 33 *Hypatia* 256, 256-257.

⁸⁵ Ibid 261. The CoE defines sexism as 'any supposition, belief, assertion, gesture or act that is aimed at expressing contempt towards a person, based on her or his sex or gender, or to consider that person as inferior or essentially reduced to her or his sexual dimension' (Council of Europe, *Gender Equality Strategy (2014-2017): Combating Sexist Hate Speech*, 2016, 2).

contrast, misogynistic speech goes beyond expression of patriarchal beliefs: it can be defined as language that directs implicit or explicit hostility towards women because they are perceived to subvert patriarchal norms.⁸⁶

The fundamental distinction can therefore be seen to lie in the components of hostility and coercion, often manifesting as harassment, intimidation or ridicule, that the latter reflects – usually through the use of aggressive language. As Richardson-Self notes, misogynistic speech ‘is not concerned with shaping people’s beliefs, theories, and values – only with compliance’.⁸⁷

Following this logic, it can be argued that sexist speech can be discriminatory without necessarily amounting to the threshold of hate speech, whereas misogynistic speech will typically constitute hate speech due to the inherent hostility it conveys. The following example illustrates this point: describing a female politician as a ‘girl’ or claiming that ‘women are too emotional for leadership’, when intended to imply incompetence or lack of qualification, and absent further aggravating context, would generally be recognised as a sexist remark, but would rarely meet the threshold for hate speech.⁸⁸ These remarks are directed at women qua group and reflect prejudice; however, it is not clear that they install hostility towards the female sex per se.⁸⁹ Conversely, if such remarks are accompanied by aggressive language – such as gendered slurs (e.g., ‘bitch’, ‘whore’, or ‘slut’) – or are compounded by gendered attacks on personal attributes that incite violence or advocate for the exclusion of the targeted individual from the political sphere, they could, upon contextual analysis, meet the threshold for hate speech.

In defining conceptual boundaries, three important points emerge concerning the identification of sexist hate speech.

First, the label ‘hate’ in ‘hate speech’ may sometimes be somewhat misleading, as it oversimplifies a range of underlying motivations.⁹⁰ Actions labelled as ‘hate speech’ are not always driven by hatred per se, but rather by various forms of prejudice, bias, or

⁸⁶ Richardson-Self (n 84) 261.

⁸⁷ Ibid.

⁸⁸ Ibid 262.

⁸⁹ Ibid.

⁹⁰ Nina Peršak, ‘Criminalising Hate Crime and Hate Speech at EU Level: Extending the List of Eurocrimes under Article 83(1) TFEU’ (2022) 33 *Criminal Law Forum* 85, 91.

intolerance, which may stem from different emotional or even non-emotional sources. Consequently, scholarly literature – particularly that grounded in psychological research – has increasingly favoured terms such as ‘prejudice’, ‘bias’ or ‘intolerance’ motivated expression as more accurate descriptors. This shift is also reflected in the gradual evolution of official definitions and legal frameworks.⁹¹

Second, an act or remark can be classified as misogynistic or hateful even if it does not explicitly target all women as a collective group.⁹² This underscores the significance of individual experiences in understanding misogyny. Rather than being inherently universal, misogyny operates primarily as a mechanism to enforce women's compliance with patriarchal gender norms. It is here pertinent to cite philosopher Kate Manne:

‘Why would any given man in a typical patriarchal setting have a problem with women universally, or even very generally, regardless of their relations? On the contrary, we would expect even the [least-]enlightened man to be well-pleased with some women, that is, those who amicably serve his interests’.⁹³

Indeed, in contrast to sexist speech – which typically involves discrimination between ‘men’ and ‘women’ as distinct social groups – misogyny often functions through a differentiation between ‘good’ and ‘bad’ women, a dynamic referred to as intradivisional misogynistic speech.⁹⁴ In this context, Richardson-Self distinguishes two types of misogynistic speech:

- (i) Speech that targets women indiscriminately, which would qualify as interdivisional misogyny, and
- (ii) Speech that selectively vilifies a specific group of women – those viewed as transgressive of patriarchal norms –, which would qualify as intradivisional misogyny.

⁹¹ Ibid.

⁹² Richardson-Self (n 84) 263.

⁹³ Richardson-Self (n 84) 263, citing Kate Manne, *Down Girl: The Logic of Misogyny* (Oxford University Press 2017) 79-80.

⁹⁴ Richardson-Self (n 84) 257.

For example, a statement such as ‘I hate women, they’re all sluts’⁹⁵ exemplifies well interdivisional misogyny, while ‘all feminists should be gang raped’⁹⁶ illustrates intradivisional misogyny.

On the one hand, while interdivisional expressions clearly constitute hate speech due to their generalised animosity toward an identified group, the classification of intradivisional forms as hate speech may be less immediately apparent, as they might more easily be disguised as legitimate political criticism directed at the individual, perhaps requiring a more careful contextual analysis. Nevertheless, when identified, the targeting of ‘only’ a subset of an oppressed group should not, in itself, justify rejecting the discourse as hate speech. As Richardson-Self argues, other women who encounter such expressions often recognise that they could just as easily become the next target, and that the previous victim was targeted precisely because she was a woman – the underlying message is, then, one of collective vulnerability and the resultant harm is widespread.⁹⁷

On the other hand, the standard examples cited above are highly explicit; in contrast, some misogynistic attacks manifest through more implicit and subtle forms of gendered targeting. As previously demonstrated in the Finnish case (see table 2), remarks instructing a female politician to ‘go and bake something or load the washing machine’ when accompanied by hostile language such as ‘empty-headed left winger’ (as in the original comment) or by gendered slurs with no male equivalent should be unequivocally recognised as not only reinforcing harmful gender stereotypes, but also inciting hostility towards women qua group. Therefore, these elements combined should be considered sufficient to bring the speech over the threshold into gender-based hate speech.

Third, as boundaries are sometimes difficult to distinguish and there are overlaps, it is important to acknowledge that whenever sexism is used in service of misogynist ends in practice, it should be also classified as hate speech.⁹⁸ While it is true that hate speech is

⁹⁵ Recounted by a twenty-four-year-old white woman (Richardson-Self, n 84, citing Laura Beth Nielsen, ‘Situating Legal Consciousness: Experiences and Attitudes of Ordinary Citizens about Law and Street Harassment’, 2000, 34, 4, *Law and Society Review*).

⁹⁶ Richardson-Self (n 84) 264.

⁹⁷ *Ibid* 266.

⁹⁸ Katarzyna Sękowska-Kozłowska, Grażyna Baranowska and Aleksandra Gliszczynska Grabias, ‘Sexist Hate Speech and the International Human Rights Law: Towards Legal Recognition of the Phenomenon by

frequently characterised by aggressive and hostile language, with its impact often shaped by these elements, such explicit expression is not a prerequisite. Gendered hate speech can take – and often does – more subtle and restrained forms, appearing moderate, unemotional, or even naively benign.⁹⁹ It may be conveyed through ambiguous humor, insinuations, imagery, or even coded language (the so-called ‘dog whistles’), making its hateful message less immediately apparent but no less harmful.¹⁰⁰ In such cases, it is particularly crucial to closely consider the context in which the speech occurs in order to avoid overregulating freedom of expression on online environments. At the same time, it is important to allow for an appropriate response when, upon closer examination, the speech is found to serve the purpose of conveying hostility or intimidating based on gender. The first comment cited in the Finnish case is a good example of less overt denigration: ‘There’s nothing more dangerous for the security threat facing Finland than the members of the government being socialist-green feminist girls with no life experience, no backbone and no understanding of security’. This statement, though presented as political criticism and appearing moderate in language, draws on gendered stereotypes to undermine the legitimacy and competence of the targeted female leaders. The use of the term ‘girls’ diminishes their authority, while the accusations of lacking ‘backbone’ and ‘understanding’ appeal to long-standing sexist tropes about women being emotionally or intellectually unfit for leadership. When understood in context – as part of a broader pattern of gendered denigration – it becomes clear that the sexist comment serves a purpose conveying hostility towards the female sex in positions of power. By the standards outlined above, it arguably constitutes hate speech. Still, it is important to highlight that the boundaries between sexist expression and hate speech – especially in terms of legal implications – remain highly contested and far from settled. Thus, any casuistic analysis must carefully balance the necessity of restricting such speech against the risk of a harmful chilling effect from excessive restriction.

the United Nations and the Council of Europe' (2022) 35(2) *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique* 2323, 2328.

⁹⁹ Bhikhu Parekh, ‘Is There a Case for Banning Hate Speech?’ in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech* (Cambridge University Press 2012) 37, 41.

¹⁰⁰ *Ibid.*

Legal implications

The legal distinction between sexist speech and sexist hate speech carries important implications for accountability and protection under national and European legal frameworks. Such a distinction is essential to uphold the principle of proportionality in legal responses and to protect the 'safety net' of freedom of expression.¹⁰¹ Reflecting this approach, GREVIO emphasises the importance of encouraging national media to adopt and enforce self-regulatory standards aimed at preventing the dissemination of content that 'feeds into hate and/or sexist speech against women'.¹⁰²

The need for distinction does not mean, however, that sexist speech shouldn't be legally sanctioned.¹⁰³ It rather means that the reasons for doing so are not necessarily the same as those that apply to hate speech.

Depending on jurisdiction, some forms of gendered hate speech, may trigger criminal or civil liability and additionally determine state obligations to investigate and sanction under international or European human rights treaties or mandate digital platforms obligations to remove content.¹⁰⁴

Notwithstanding the multiple references to the issue, when employing the term 'sexist hate speech' in official documents, the CoE does not provide a precise legal definition, nor does it delineate the specific legal parameters of such behaviour. However, CoE's Recommendation on Combating Hate Speech explicitly mandates States to 'specify and clearly define in their national criminal law which expressions of hate speech are subject to criminal liability' and provides concrete examples to guide implementation. With regard to the gender dimension, the Recommendation identifies specific forms of sexist hate speech that should be criminalised, including: (i) Public incitement to violence or

¹⁰¹ Ibid.

¹⁰² Council of Europe, GREVIO, *Baseline Evaluation Report Andorra* (2020), para 88; Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2338-39.

¹⁰³ In the 1950s Gordon Allport was already conceptualising a 'pyramide of hate' or 'scale of prejudice' highlighting that offensive speech, even in its legal and protected form, may contribute to increasing more severe forms of speech and, indeed, actions against protected groups (European Union Agency for Fundamental Rights, *Online Content Moderation – Current Challenges in Detecting Hate Speech*, FRA, 2023, 21, citing Gordon Allport, *The Nature of Prejudice*, Addison-Wesley 1954).

¹⁰⁴ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, paras 2-3, 11, 16-18; Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2330-31.

discrimination; (ii) Sexist threats; and (iii) sexist insults – particularly under conditions outlined for online insults in the Additional Protocol to the Convention on Cybercrime (Budapest Convention),¹⁰⁵ and (iv) Intentional dissemination of material containing any of the above expressions of hate speech.¹⁰⁶

This further underscores that even within the category of hate speech itself, there are varying degrees of severity, and accordingly, the sanctions imposed must reflect a proportionate gradation. While all hate speech should be deemed unlawful, not all instances warrant criminalization.

At the same time, sexist speech will typically be protected under the umbrella of freedom of expression,¹⁰⁷ although at the national level it might still be sanctioned under anti-discrimination frameworks, particularly when it amounts to harassment or acquires a systematic nature.¹⁰⁸

This legal differentiation highlights the complexity of regulating gender-based discriminatory expression, particularly in balancing the right to freedom of expression with the obligation to protect against discrimination, gender-based violence, and hate speech¹⁰⁹ - a complexity that will be further examined in Chapter III through the lens of the relevant legal frameworks.

2.2.2. Gendered Disinformation

¹⁰⁵ See Articles 1 to 7 of the Protocol to the Convention.

¹⁰⁶ Council of Europe, *Recommendation CM/Rec(2022)16 of the Committee of Ministers to member States on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, para 11.

¹⁰⁷ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2330-32.

¹⁰⁸ Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019)1 including Appendix; Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2329-30. Regarding cyber harassment, see, for example, Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence [2024] OJ L 1385/1, Art 7.

¹⁰⁹ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, paras 1, 6; Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2331-32.

Alongside hate speech, gendered disinformation has emerged as a particularly alarming dimension of gendered abuse directed at public figures.¹¹⁰ Despite the growing recognition, there remains no clear consensus on how to define this broader, evolving phenomenon. As Special Rapporteur Irene Khan has noted, this lack of definitional clarity is compounded by the absence of contextualised research and disaggregated data – critical tools for capturing the multifaceted and complex nature of gendered disinformation.¹¹¹ These gaps continue to obstruct efforts to fully understand its distinct characteristics and far-reaching impact.¹¹²

Nevertheless, certain defining traits can already be identified. A crucial starting point is distinguishing between disinformation and misinformation, given that these two concepts are frequently referred to collectively as 'fake news', despite important differences between them. The key distinction lies in intent. Misinformation is the dissemination of false information without the intention of causing harm, albeit harm may ultimately result.¹¹³ Conversely, disinformation is deliberately created and spread with the specific aim of causing harm to individuals, organisations or nations, and of manipulating public opinion.¹¹⁴ Notably, malicious intent and the potential to cause public harm are key defining elements. Accordingly, this definition excludes accurate information, unintentional inaccuracies, satire, parody, irony, and other forms of expression such as humour or commentary that are not intended to mislead.¹¹⁵

Although a universal conceptualisation of 'gendered disinformation' has yet to be achieved, the phenomenon has been identified as involving the use of 'false or misleading gender and sex-based narratives against women, often with some degree of coordination, aimed at deterring women from participating in the public sphere'.¹¹⁶ Further elaborating the means through which it operates, the think tank Demos defines it as 'information

¹¹⁰ Irene Khan, *Report of the Special Rapporteur on Disinformation and Freedom of Opinion and Expression* (13 April 2021) UN Doc A/HRC/47/25, para 27.

¹¹¹ *Ibid* para 28.

¹¹² *Ibid*.

¹¹³ Publications Office of the European Union, *Tackling Disinformation and Information Manipulation* (Fact Sheet, June 2024), 1.

¹¹⁴ *Ibid*.

¹¹⁵ Alexander Lanoszka, 'Disinformation in International Politics' (2019) 4 *European Journal of International Security* 227, 229.

¹¹⁶ Nina Jankowicz and others, 'Malign Creativity: How Gender, Sex and Lies Are Weaponized Against Women Online' (Wilson Center, Science and Technology Innovation Program, January 2021) 1.

activities (creating, sharing, disseminating content) which attack or undermine people on the basis of their gender; weaponise gendered narratives to advance political, social, or economic objectives.’¹¹⁷

As with all manifestations of OVAW, this tactic is sustained in deep rooted gender biases that portray women as unfit for leadership, based on assumptions that they are inherently ‘unworthy’ (often perceived as dishonest), ‘unqualified’, ‘unintelligent’ or ‘unlikeable’.¹¹⁸ When employed in the context of politics, at its core, gendered disinformation weaponizes rumours and stereotypes, spreading false, deceptive, or hateful narratives, often laced with abusive language, to undermine female candidates and influence political outcome.¹¹⁹

This phenomenon, as delineated by prevailing definitions of ‘disinformation’, will encompass three defining characteristics: (i) falsity, (ii) malign intent and (iii) some degree of coordination.¹²⁰

While the concepts of falsity and malign intent are relatively straightforward, the element of coordination can be more difficult to detect, as it often manifests in covert, less obvious, forms. In this regard, sociologist Sarah Sobieraj identifies three overlapping strategies used by aggressors to silence women or to limit their impact in the digital publics:

- (i) ‘Intimidating’, which commonly involves threats to women’s physical integrity (e.g. rape or death threats), but also other cyberattacks such as Distributed Denial-of-Service (DDoS),¹²¹ spamming,¹²² or doxing;

¹¹⁷ Ellen Judson and others, *Engendering Hate: The Contours of State-Aligned Gendered Disinformation Online* (Demos 2020) 7.

¹¹⁸ Kristina Wilfore, ‘Security, Misogyny, and Disinformation Undermining Women’s Leadership’ in Gulizar Hacıyakupoglu and Yasmine Wong (ed), *Gender and Security in the Digital Space* (Routledge 2022) 124, 131.

¹¹⁹ Judson and others (n 117) 11.

¹²⁰ Wilfore (n 118) 131.

¹²¹ DDoS Attack is a cybercrime in which the attacker floods a server with internet traffic to prevent users from accessing connected online services and sites (Fortinet, ‘DDoS Attack’, *Fortinet*, 2024, <https://www.fortinet.com/resources/cyberglossary/ddos-attack> accessed 27 April 2025).

¹²² Spamming, as defined by the Cambridge Business English Dictionary, is the activity of sending advertisements by email to people who do not want to receive them (Cambridge Dictionary, ‘Spamming’, Cambridge University Press, <https://dictionary.cambridge.org/dictionary/english/spamming> accessed 27 April 2025).

- (ii) ‘Shaming’, which aims to exploit societal prejudices on women’s sexuality and appearance, often through unauthorised release of private information or images and the dissemination of false and humiliating narratives;
- (iii) ‘Discrediting’, which involves efforts to undermine women’s credibility seeking to portray them as unfit for elected office, delegitimising their opinions and contributions to the public discourse.¹²³

The element of coordination can also be an indicator of the so-called ‘state-aligned’ disinformation practices, which have predominantly been observed in authoritarian states.¹²⁴ These practices are broadly understood to be the dissemination of disinformation not only by state-affiliated actors, but also by informal or proxy groups that use gendered narratives to shield the state from criticism or perceived threats to its legitimacy, particularly when such criticism originates from women in public or political roles.¹²⁵ Furthermore, Kristina Wilfore argues that disinformation systematically targeting female politicians can be a deliberate tactic to undermine democratic institutions, and as such, it must not be neglected in analyses of antidemocratic trends.¹²⁶

The perceived ‘neutrality’ of information on social media contributes to reinforcing internal biases against women, while sensational headlines – such as ‘slept her way to the top’ – as well as the ‘click rate’ effect accelerate the spread of sexist and misogynistic content on social media.¹²⁷ In this ecosystem, it is also essential to recognise that although public debates on ‘fake news’ have largely focused on text-based content, a substantial portion of online disinformation now circulates through memes¹²⁸ and audiovisual

¹²³ Jankowicz and others (n 116) 8; Sarah Sobieraj, ‘Bitch, Slut, Skank, Cunt: Patterned Resistance to Women’s Visibility in Digital Publics’ (2018) 21(11) *Information, Communication & Society* 1700, 1704-1707.

¹²⁴ Judson and others (n 117) 7.

¹²⁵ Ibid.

¹²⁶ Wilfore (n 118) 125. See also Irene Khan, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (7 August 2023) UN Doc A/78/288, para 18.

¹²⁷ For instance, studies have showed that social media users are easily moved by ‘emotion’ rather than sustained evidence (Xiangyu Ouyang and others, ‘Disinformation Reinforces Female Political Inequality and Social Misogyny’ in *Proceedings of the 2021 International Conference on Public Relations and Social Sciences*, Atlantis Press, Advances in Social Science, Education and Humanities Research, vol 586, 2021, 247, 249).

¹²⁸ A ‘meme’ can be understood as a collection of multimodal digital content – created, transformed, and circulated online – which is context/culture-inspired and whose composition is technically shaped by the

formats.¹²⁹ Often informal and easily shareable, these formats are particularly effective at embedding harmful gendered narratives in public discourse.

The 2024 European Parliamentary elections offer a relevant perspective on the scope and scale of the phenomenon. In this context, the Global Disinformation Index recorded a significant increase in online abuse and disinformation targeting female EU leaders and candidates.¹³⁰ These attacks frequently combined misogynistic rhetoric with conspiracy theories and xenophobic narratives.¹³¹

At national level, three scenarios reinforce this pattern. In Germany, research by the Institute for Strategic Dialogue revealed a marked disparity in the spread of false or conspiratorial content targeting female candidate Annalena Baerbock, compared to her male opponents Olaf Scholz and Armin Laschet. Notably, these attacks were often overtly gendered, focusing on Baerbock's appearance and her alleged incompetence, rather than her political agenda.¹³²

In Croatia, former President Kolinda Grabar-Kitarović was targeted by a sustained campaign that hypersexualised her image, ultimately aiming to sway the outcome of the 2020 presidential elections.¹³³

In Ukraine, MP Svitlana Zalishchuk became the target of gendered disinformation after her widely praised UN speech on the war's impact on women in 2017. A fake tweet falsely attributed to her – claiming she would run naked through Kyiv if the Ukrainian

platforms where it appears (Richard Rogers and Giulia Giorgi, 'What is a Meme, Technically Speaking?', 2024, 27 *Information, Communication & Society* 73; Limor Shifman, *Memes in Digital Culture*, MIT Press 2014).

¹²⁹ International Media Support (IMS), *Online Gendered Disinformation and Sexist Hate Speech, Intersectional Feminist Media Development Learning Brief No 2*, 1, 3. For instance, in France and the United Kingdom, according to the election-related projects First Draft, visual content circulating on Facebook proved to be both the most widely disseminated and the most challenging to verify or debunk among misleading materials (Matteo Moschella and Ryan Watts, 'What We Learned Fact-Checking the UK Election', *First Draft*, 19 June 2017, <https://firstdraftnews.com/joint-venture-learnings/> accessed 26 April 2025).

¹³⁰ Global Disinformation Index, 'Gendered Disinformation in the European Parliamentary Elections' (10 June 2024) <https://www.disinformationindex.org/blog/2024-06-10-gendered-disinformation-in-the-european-parliamentary-elections/> accessed 6 March 2025.

¹³¹ Ibid.

¹³² Wilfore (n 118) 127.

¹³³ Ibid 130; Emma Goldberg, 'Fake Nudes and Real Threats: How Online Abuse Holds Back Women in Politics' *The New York Times* (4 June 2021) <https://www.nytimes.com/2021/06/03/us/disinformation-online-attacks-female-politicians.html> accessed 6 May 2025.

army lost a key battle – was circulated alongside doctored nude images, clearly intended to humiliate and discredit her.¹³⁴

Coincidence or not, neither Grabar-Kitarović nor Zališčuk were re-elected in the elections that followed.

The facilitated spread of disinformation on social media can significantly shape public perceptions of female candidates, fuelling misunderstandings and intensifying hostility from political opponents.¹³⁵ For supporters, such false narratives may erode trust and diminish backing for the targeted candidates; for those who already hold gender-based stereotypes, they may serve to reinforce existing biases by presenting a confirmation of their views through a distorted portrayal.¹³⁶

Gendered Disinformation as Gender-Based Violence

Despite the preceding discussion, it has been argued that overly broadening the definition of gendered disinformation risks conflating it with the concept of gender-based violence, which still warrants distinction.¹³⁷

In her 2023 UN Report on gendered disinformation and its implications for freedom of expression, Special Rapporteur Irene Khan observed that the concepts of gender-based violence, gendered hate speech, and gendered disinformation, although interrelated, require careful conceptual distinction.¹³⁸ As Khan explains, the persistent conflation of these phenomena can be attributed, first, to the lack of consensus on how to define gendered disinformation or on the most effective strategies for combating it; second, the ways in which the phenomenon contributes to online gender-based violence, and how it

¹³⁴ Nina Jankowicz, ‘How disinformation became a new threat to women’ *Coda Story* (11 December 2017) <https://www.codastory.com/polarization/how-disinformation-became-a-new-threat-to-women/> accessed 17 June 2025.

¹³⁵ Xiangyu Ouyang and others, ‘Disinformation Reinforces Female Political Inequality and Social Misogyny’ in *Proceedings of the 2021 International Conference on Public Relations and Social Sciences* (Atlantis Press, Advances in Social Science, Education and Humanities Research, vol 586, 2021) 247, 249.

¹³⁶ *Ibid.*

¹³⁷ Bardall (n 71) 112.

¹³⁸ Irene Khan, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (7 August 2023) UN Doc A/78/288, para 2.

intersects with broader social inequalities, are still poorly understood, underscoring the complexity and inherently political dimensions of the issue.¹³⁹

In line with this reasoning, Bardall argues, that the core issue in the phenomenon of gendered disinformation is not simply the expression of biased opinions, but the intentional and deceptive manipulation of information – such as fabrications, rumours, and manipulated images – used to reinforce harmful gender stereotypes. In this regard, the author states that the term gendered disinformation calls for differentiation in relation to gender-based violence, as it does not introduce new biases but rather amplifies existing ones.¹⁴⁰

Although I acknowledge the need to define these issues separately, I would argue that gendered disinformation targeting female politicians is, in fact, a form of gender-based violence.¹⁴¹

On the one hand, such campaigns align with the core characteristics of VAWIP, as a recognised form of gender-based violence. They target women specifically because they are women or affect them disproportionately compared to male counterparts in similar positions, they are often gendered in form – through, for example, the use of sexualised imagery or gendered slurs – and they can have a chilling effect on women's political participation by manipulating public perception and causing psychological harm to the individual.

On the other hand, in my view, the intentional manipulation of information through the creation of deceptive content does more than merely reproduce or amplify existing gender biases – it actively reinforces and entrenches them. Moreover, gendered disinformation encompasses not only the act of creation but also intentional sharing and disseminating.¹⁴² Drawing a parallel with hate speech – where the intentional dissemination of hateful content is considered as punishable as its original expression¹⁴³ – there appears to be no

¹³⁹ Ibid.

¹⁴⁰ Bardall (n 71) 112.

¹⁴¹ For a supporting perspective see, for example, Wilfore (n 118) 126.

¹⁴² Intent can be inferred, for example, from a suggestive caption accompanying the shared content.

¹⁴³ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, para 11.

reasonable justification for treating the deliberate dissemination of disinformation as less serious than its initial creation.

In sum, and prior to the legal analysis in section 3.3., gendered disinformation should be understood as a hybrid phenomenon: it operates as a form of information disorder by distorting public discourse and threatening democratic integrity, while simultaneously embodying a gendered dimension that necessitates protection through specialised legal frameworks.

Chapter III: Freedom of expression v. Hate Speech: Where Does Gendered Discourse Fit Into the Debate?

Following a preliminary conceptualisation of gendered hate speech and gendered disinformation, I now turn to their analysis in light of the relevant legal frameworks in order to situate both phenomena within the ‘freedom of expression v. hate speech’ debate.

Before turning to the substance of this Chapter, two introductory remarks must be made regarding the definition of hate speech under international standards.

Firstly, no universally accepted definition of hate speech can be found in either international or European human rights law, as ongoing debates continue to surround its meaning and scope.¹⁴⁴ Attempts to define it can be seen in soft-law instruments, which reflect diverse approaches over UN, CoE and EU frameworks, and tend to offer guiding principles rather than precise legal definitions.

Secondly, given this lack of universality, a longstanding concern has been the dilution of the concept, leading to the overregulation of speech, including online, and, consequently, to excessive restrictions that in some cases may reflect some form of state censorship.¹⁴⁵

¹⁴⁴ United Nations, *United Nations Strategy and Plan of Action on Hate Speech* (2019) 2.

¹⁴⁵ Maria Sjöholm, 'Regulation of Online Gender-Based Hate Speech and International Human Rights Law: Current Status and Challenges' (2024) 104 *Questions of International Law* 3, 4.

For instance, the issue arose during the negotiations of CEDAW, where, reflection of the historical context of the time, a proposal to include in the text of the treaty the need to eliminate ‘sexist hate speech’ was ultimately rejected.¹⁴⁶

One objection to restricting gender-based insults under hate speech laws is that such language is so deeply embedded in everyday discourse that prohibiting it would result in an extensive limitation on speech, potentially leading to excessive censorship and a significant curtailment of free expression.¹⁴⁷ Nevertheless, this concern is not decisive. Historically, derogatory language related to race and religion was similarly pervasive, yet this ubiquity was not accepted as a valid reason to avoid challenging or regulating such speech.¹⁴⁸ As legal scholar Alexander Brown emphasises, the fact that a law might restrict a considerable amount of routine behaviour is not, in itself, a sufficient argument against its legitimacy.¹⁴⁹

This recognition, however, does not negate the need for a nuanced and cautious approach to hate speech regulation. While advocating for stronger legal protections against hate speech, it remains essential to distinguish such speech as a particularly grave and narrowly defined category, requiring careful contextual interpretation when assessing each case.¹⁵⁰

That being said, while free speech is the indispensable basis of free thought and critical self-consciousness – both vital to political life –, it is not the only key value that shapes democratic societies.¹⁵¹ As political theorist Bhikhu Parekh argues, human dignity,

¹⁴⁶ Ibid; Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2232-33.

¹⁴⁷ Alexander Brown, ‘The Who Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches’ (2017) 30 Canadian Journal of Law & Jurisprudence 23, 54. Free speech absolutists also warn that excessive regulation may unintentionally harm the very groups it aims to protect by promoting paternalism and reinforcing institutional notions of victimhood. (George Wright, ‘Dignity and Conflicts of Constitutional Values: The Case of Free Speech and Equal Protection’, 2006, 43 *San Diego Law Review* 527, 566). This concern is echoed by the UN Special Rapporteur on Freedom of Expression, who has criticised efforts to regulate online content in the name of safeguarding cultural or moral values as inherently paternalistic (Abid Hussain, ‘Report of the Special Rapporteur Submitted Pursuant to Commission on Human Rights Resolution 1997/26’, 28 January 1998, UN Doc E/CN.4/1998/40, para 45).

¹⁴⁸ Alexander Brown, ‘The Who Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches’ (2017) 30 Canadian Journal of Law & Jurisprudence 23, 54.

¹⁴⁹ Ibid.

¹⁵⁰ Parekh (n 99) 40; UN Human Rights Committee, *General Comment No 34: Article 19: Freedoms of Opinion and Expression* (12 September 2011) UN Doc CCPR/C/GC/34, para 11. See, in this respect, the distinction made between ‘sexist speech’ and ‘sexist hate speech’ in section 2.2.1.

¹⁵¹ Parekh (n 99) 40.

equality, the right to live free from harassment and intimidation, and mutual respect are just as important and, thus, deserve equivalent level of protection by law.¹⁵²

Therefore, freedom of expression – though protected to varying degrees depending on each jurisdiction’s historical context – is not an absolute right in any international framework. It can and should be subjected to restrictions under specific and justified circumstances prescribed by law, where striking a legal balance based on proportionality becomes central.

3.1. The Council of Europe

Context and Definition

In clear contrast to other approaches to freedom of expression – paradigmatically, the one taken by the United States, bound by the First Amendment to their Constitution¹⁵³ –, the CoE allows for limitations under specific circumstances. In fact, over the past two decades, it has significantly broadened the scope of expression considered impermissible, going beyond the prohibition of racist hate speech or genocide denial. Notably, the European Court of Human Rights (hereafter ECtHR or ‘the Court’) has explicitly recognised homophobic hate speech as a prohibited form of speech in several landmark judgements.¹⁵⁴

Regarding sexist hate speech, while the ECtHR has acknowledged its existence, it has largely done so at the margins, with no currently established body of jurisprudence specifically addressing the issue.¹⁵⁵ Nevertheless, certain developments in the Court’s

¹⁵² Ibid 43.

¹⁵³ ‘Congress shall make no law [...] abridging the freedom of speech, or of the press [...]’ (United States Congress, *Amendment I: Freedom of Religion, Speech, Press, Assembly, and Petition*, Constitution Annotated, <https://constitution.congress.gov/browse/amendment-1/> accessed 2 May 2025).

¹⁵⁴ See *Vejdeland and Others v Sweden* App no 1813/07 (ECtHR, 9 February 2012), where the Court recognised sexual-orientation-based hate speech for the first time, and *Lilliendahl v Iceland* App no 29297/18 (ECtHR, 12 May 2020), which in para 39 confirmed that offensive statements targeting homosexuals could fall outside the protection of Article 10 of the European Convention on Human Rights.

¹⁵⁵ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2337.

case-law, such as those on homophobic hate speech, offer valuable insights that can be extended to the gendered form of the phenomenon.¹⁵⁶

The CoE first addressed hate speech as a concept in a 1997 Recommendation, defining it as incitement to hatred directed at specific individuals or groups with certain protected characteristics. In this regard, it included ‘all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance [...]’.¹⁵⁷ In this definition the omission of ‘sex’ or ‘gender’ as a protected characteristic – among many others – is notorious.¹⁵⁸

Nevertheless, the CoE Commissioner for Human Rights defended this choice by emphasising the open-ended character of the 1997 definition which in ‘other forms’ should be understood as non-exhaustive and implicitly inclusive of other vulnerable groups, including women and girls.¹⁵⁹

In 2015, the European Commission against Racism and Intolerance (ECRI) expanded the definition of hate speech to include:

‘[...] the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of "race", colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status.’¹⁶⁰

By incorporating terms such as ‘harassment, insult, negative stereotyping, stigmatization, and threat’, and by significantly broadening the range of protected grounds, this

¹⁵⁶ Ibid.

¹⁵⁷ Council of Europe, *Recommendation No R (97) 20 of the Committee of Ministers to Member States on 'Hate Speech'* (adopted 30 October 1997).

¹⁵⁸ An explanatory memorandum was issued alongside the recommendation, clarifying that this absence was a deliberate choice aimed at maintaining the document’s thematic clarity. The memorandum notes that issues of intolerance based on sex had previously been addressed in the CoE’s Committee of Ministers’ 1984 Recommendation concerning gender equality in the media (Council of Europe, *Recommendation No R (84) 17 of the Committee of Ministers to Member States on Equality Between Women and Men in the Media*, adopted 25 September 1984). However, this earlier recommendation did not pertain to hate speech and hence could not be interpreted as covering sexist hate speech.

¹⁵⁹ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2335.

¹⁶⁰ European Commission against Racism and Intolerance (ECRI), *General Policy Recommendation No 15 on Combating Hate Speech* (Council of Europe, 2015) 3.

development marked a substantial step forward in standard-setting. It enables a more nuanced and inclusive identification of gender-based hate speech, which often manifests through these less overt forms, as discussed in section 2.2.1. of this thesis.

In this context, despite the Parliamentary Assembly of the CoE (PACE) urging in Recommendation No. 2098 (2017)¹⁶¹ and Resolution No. 2144 (2017)¹⁶² for a broader definition by the Committee of Ministers, only in 2022 was the 1997 definition expanded to explicitly include protected grounds other than race, ethnicity or national origin – notably, ‘sex’ was one of them.¹⁶³ However, it remained narrower in scope than ECRI’s definition as it excluded expressions such as ‘harassment, insult, negative stereotyping, stigmatisation or threat’. This ultimately hinders the identification of important dimensions of gender-based hate speech by the CoE’s central decision-making body, which reflects the consensus of the 46 member states.

Nonetheless, a coherent set of European standards can still be discerned with reasonable clarity through the evolution of national laws and practices. Across jurisdictions, hate speech is generally understood as public expression that incites hatred against individuals or groups based on arbitrary and immutable identity traits, like sex. This type of speech:

- (i) Directly endangers the rights and safety of targeted individuals and communities;
- (ii) Undermines core societal values such as equality, dignity, pluralism, and inclusion, thereby eroding the foundations of democratic societies;
- (iii) Ultimately contradicts the very principles on which freedom of expression is grounded, calling into question its legitimacy when used to deny others their fundamental rights.¹⁶⁴

¹⁶¹Parliamentary Assembly of the Council of Europe, *Recommendation 2098 (2017) on ending cyberdiscrimination and online hate* (25 January 2017) Rec 2098 (2017).

¹⁶²Parliamentary Assembly of the Council of Europe, *Resolution 2144 (2017) Cyberdiscrimination and Hate Speech on the Internet* (25 January 2017) Res 2144 (2017) paras 2 and 7.2.3.

¹⁶³Committee of Ministers, *Explanatory Memorandum to Recommendation CM/Rec(2022)16* (20 May 2022) Appendix para 2.

¹⁶⁴Julia Kapelańska-Pręgowska and Maja Pucelj, ‘Freedom of Expression and Hate Speech: Human Rights Standards and Their Application in Poland and Slovenia’ (2023) 12 *Laws* 1, 4; Cvijeta Senta and others, *Preventing Hate Speech Online: Materials for Teachers* (Centre for Peace Studies Zagreb, 2021) 32.

The European Convention on Human Rights

Turning to the analysis of the European Convention on Human Rights (hereafter ECHR or ‘the Convention’), Article 10(1) serves as the cornerstone of the protection of freedom of expression in the European context. It provides that ‘everyone has the right to freedom of expression’, including the ‘freedom to hold opinions and to receive and impart information and without interference by public authority and regardless of frontiers. [...]’. Nonetheless, Article 10(2) ECHR refers to duties and responsibilities engaged by the exercise of freedom of expression as a non-absolute right.

In this context, hate speech will typically fall outside the scope of protection afforded by Article 10(1) ECHR. The concept and defining elements of hate speech have been progressively shaped through the Court’s jurisprudence, beginning with their initial formulation in the 1999 judgment of *Sürek v. Turkey (No. 4)*.¹⁶⁵

In the absence of a legally binding definition, the ECtHR’s approach to hate speech remains highly case specific. This approach has evolved into a two-tiered framework, distinguishing between more severe forms – typically addressed under Article 17 ECHR (prohibition of abuse of rights) – and less severe forms, which fall within the scope of Article 10(2) (permissible restrictions on freedom of expression).

On the one hand, the ECtHR typically invokes Article 17 in cases where speech directly incites violence or hatred, or seeks to undermine the rights of others. The Court first clarified the function of Article 17 in *Lawless v. Ireland* (1961), explaining that it prevents individuals or groups from using the Convention to justify actions aimed at ‘destroying’ the rights and freedoms it guarantees.¹⁶⁶ Where Article 17 applies, such speech is excluded a priori from the protection of Article 10(1).¹⁶⁷

¹⁶⁵ *Sürek v Turkey* (No 4) App no 24762/94 (ECtHR, 8 July 1999) para 60; Julia Kapelańska-Pręgoszka and Maja Pucelj, ‘Freedom of Expression and Hate Speech: Human Rights Standards and Their Application in Poland and Slovenia’ (2023) 12 *Laws* 1, 4.

¹⁶⁶ *Lawless v Ireland* App no 332/57 (ECtHR, 1 July 1961) para 7. Rather than applying it directly, the ECtHR often employs Article 17 as an interpretative tool to define the boundaries of other Convention rights – most notably, Article 10 on freedom of expression. (William A Schabas, *The European Convention on Human Rights: A Commentary*, OUP 2015, 614-616).

¹⁶⁷ Julia Kapelańska-Pręgoszka and Maja Pucelj, ‘Freedom of Expression and Hate Speech: Human Rights Standards and Their Application in Poland and Slovenia’ (2023) 12 *Laws* 1, 4.

On the other hand, where the Court proceeds under Article 10(2), there is not an a priori exclusion from the material scope of Article 10(1).¹⁶⁸ Instead, certain criteria must be met for a restriction on expression to be deemed lawful. The interference must (i) be prescribed by law, (ii) pursue one or more legitimate aims, and (iii) be necessary in a democratic society – implying that it must also be proportionate to the aim pursued. The same provision sets out a relatively extensive list of justifications for restrictions, ranging from ‘national security’ to the ‘reputation or rights of others’. In hate speech cases, the primary concern revolves around the latter category. Although when recognising a need for restriction in light of the ‘rights of others’ the Court does not always identify the specific right being violated, Article 8 (right to private life) and Article 14 (right to non-discrimination) of the Convention are the most common considerations.¹⁶⁹

The ECtHR’s assessment of whether a restriction on expression is ‘necessary in a democratic society’ involves evaluating several contextual factors, including the political and social climate,¹⁷⁰ the speaker’s intention,¹⁷¹ the nature and content of the expression,¹⁷² the speaker’s societal role or status,¹⁷³ the scope or reach of the expression,¹⁷⁴ and the characteristics of the audience addressed.¹⁷⁵

A major obstacle in addressing hate speech, both online and offline, stems from a principle the ECtHR has consistently emphasised: Article 10 ECHR protects not only expressions that are favourably received, considered inoffensive or viewed as indifferent, but also those that may provoke or unsettle, ‘such are the demands of that pluralism, tolerance and broadmindedness without which there is no “democratic society”’.¹⁷⁶

Consequently, navigating the tension between the individual’s right to voice opinions that might ‘offend, shock or disturb’ – as articulated in the ECtHR’s landmark *Handyside v.*

¹⁶⁸ See, for example, *Perinçek v Switzerland* [GC] App no 27510/08 (15 October 2015) paras 113-15.

¹⁶⁹ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, preamble.

¹⁷⁰ See, for example, *Leroy v France* App no 36109/03 (ECtHR, 2 October 2008) para 45.

¹⁷¹ See, for example, *Jersild v Denmark* App no 15890/89 (ECtHR, 9 September 1994) para 31.

¹⁷² See, for example, *Ottan v France* App no 41841/12 (ECtHR, 19 April 2018) paras 62-68.

¹⁷³ See, for example, *Incal v Turkey* App no 22678/93 (ECtHR, 9 June 1998) para 67-68.

¹⁷⁴ See, for example, *Gündüz v Turkey* App no 35071/97 (ECtHR, 04 December 2003) para 49.

¹⁷⁵ See, for example, *Vejdeland and Others v Sweden* App no 1813/07 (ECtHR, 09 February 2012) para 56.

¹⁷⁶ *Handyside v United Kingdom* App no 5493/72 (ECtHR, 7 December 1976) para 49.

United Kingdom (1976) judgment¹⁷⁷ – and the imperative to shield others from expressions of hatred constitutes a persistent dilemma for both courts and policymakers. The ECtHR extended this reasoning to the digital sphere in subsequent judgments.¹⁷⁸

As will be explored in the next section, striking this balance is especially challenging in political contexts, where the Court’s jurisprudence affords heightened protection to speech under Article 10(1) in order to safeguard democratic principles.¹⁷⁹

Nevertheless, it is possible to argue that once an instance of hate speech has been established, it will be rare – if not exceptional – for a restriction on that expression to be found unnecessary or lacking a legitimate aim under the framework of a democratic society.¹⁸⁰ In *Gündüz v. Turkey* (2003) the ECtHR affirmed that, based on the principles of tolerance and respect for the equal dignity of all human beings – fundamental pillars of a democratic and pluralistic society – it may be necessary ‘to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance [...], provided that any “formalities”, “conditions”, “restrictions” or “penalties” imposed are proportionate to the legitimate aim pursued’.¹⁸¹ The Court further stated: ‘there can be no doubt that concrete expressions constituting hate speech, which may be insulting to particular individuals or groups, are not protected by Article 10 of the Convention’.¹⁸²

Importantly, the Court has observed that incitement to hatred does not automatically imply a call for violent or other criminal actions. For instance, in *Vejdeland and Others v. Sweden* (2012), where in question was the placing of homophobic leaflets in the lockers of school children, the Court examined for the first time whether homophobic literature amounted to hate speech. In that regard, it stressed that:

‘[a]ttacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating racist speech in the face

¹⁷⁷ Ibid. See also *Observer and Guardian v United Kingdom* App no 13585/88 (ECtHR, 26 November 1991) para 59.

¹⁷⁸ See, for example, *Yıldırım v Turkey* App no 3111/10 (ECtHR, 18 December 2012) and *Cengiz and Others v Turkey* App no 48226/10 and 14027/11 (ECtHR, 1 December 2015).

¹⁷⁹ See, for example, *Lingens v Austria* App no 9815/82 (ECtHR, 8 July 1986) para 42.

¹⁸⁰ *Kapelańska-Pręgoska and Pucelj* (n 167) 5.

¹⁸¹ *Gündüz v Turkey* App no 35071/97 (ECtHR, 4 December 2003) para 40.

¹⁸² Ibid.

of freedom of expression exercised in an irresponsible manner [...]. Discrimination based on sexual orientation is as serious as discrimination based on ‘race, origin or color’.¹⁸³

Although, in this case, the Court refrained from categorising the leaflets among the most serious forms of hate speech – thus stopping short of applying Article 17 – it nevertheless found the interference to be justified under Article 10(2) ECHR.¹⁸⁴

Building on the Court’s reasoning – which equates the severity of racist and homophobic hate speech and has influenced the CoE’s working definition of hate speech – developments in the regulation of racist hate speech offer a valuable reference point for shaping current and future approaches to hate speech based on other grounds, including gender. Indeed, CoE’s Recommendation on Preventing and Combating Sexism affirms that ‘States are encouraged to take responsibility for combating hate speech and ensuring that the same rules apply to sexist hate speech as those developed for racist hate speech when it comes to the use of criminal law sanctions’.¹⁸⁵

While this alignment marks a significant step forward, it also highlights the persistent challenge of effectively identifying and addressing gender-based hate speech in practice. A fundamental distinction between the ratio behind regulation or sanctioning of racist hate speech and gendered hate speech must be acknowledged. This is because when an expression is deemed racist, it is almost universally recognised as hate speech; in contrast, identifying hate speech against women seems to be far more complex and nuanced, posing significant challenges in its classification and recognition, as discussed in the previous Chapter.¹⁸⁶ In this regard, legal scholar Mari Matsuda argues that gender-based hate speech should require ‘a separate analysis because of the complex and violent nature of gender subordination, and the different way in which sex operates as a locus of

¹⁸³ *Vejdeland and Others v Sweden* App no 1813/07 (ECtHR, 9 February 2012) para 55.

¹⁸⁴ Subsequent judgments have more explicitly addressed hate speech based on sexual orientation, affirming that such expressions fall outside the scope of protection afforded by Article 10(1). See, for example, *Lenis v Greece* App No 47833/20 (ECtHR, 27 June 2023), *Beizaras and Levickas v Lithuania* App no 41288/15 (ECtHR, 14 January 2020), *Lilliendahl v Iceland* App no 29297/18 (ECtHR, 12 May 2020), *Sabalic v Croatia* App no 50231/14 (ECtHR, 14 January 2021) and *Oganezova v Armenia* Apps nos 71367/12 and 72961/12 (17 August 2022).

¹⁸⁵ Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019)1 including Appendix, section II.B. para 2.

¹⁸⁶ *Ibid.*

oppression'.¹⁸⁷ This highlights the need to raise awareness about the specific characteristics of gender-based hate speech to ensure it is properly identified and adjudicated in judicial proceedings.

This difficulty in legal recognition is also reflected in the Court's evolving, yet cautious, approach to addressing gender-based abuse, particularly in its technology-facilitated forms. Although the Court has not explicitly recognised gender-based hate speech, it has shown growing recognition of online gender-based violence as a broader structural issue. This was first acknowledged in *Volodina v. Russia* (Nos 1 and 2 – 2019 and 2021 respectively) and *Buturugă v. Romania* (2020), in which the Court argued that states have a positive obligation to criminalise, investigate, and stop technology-facilitated violence.¹⁸⁸ However, the Court has since shown reluctance to engage with a deeper analysis of the gendered dimensions of such abuse. In *M.Ş.D. v. Romania* (2025), which concerned the non-consensual dissemination of intimate images, it recognised the gravity these conducts and underscored the shortcomings of the national legal framework in offering an effective remedy, but declined to examine the applicant's claims under Article 14 ECHR (prohibition of discrimination), finding a violation solely under Article 8 ECHR (right to respect for private life).¹⁸⁹

The Additional Protocol to the Convention on Cybercrime (2003)

¹⁸⁷ Mari J Matsuda, 'Public Response to Racist Speech: Considering the Victim's Story' (1989) 87(8) *Michigan Law Review* 2320, 2332.

¹⁸⁸ *Volodina v Russia* (No 1) App no 41261/17 (ECtHR, 4 October 2019) paras 77, 91 and 98; *Volodina v Russia* (No 2) App no 40419/19 (ECtHR, 14 December 2021) paras 48, 49 and 57; *Buturugă v Romania* App no 56867/15 (ECtHR, 11 February 2020) paras 74 and 79; Catherine Van de Heyning, 'Strasbourg's Consolidation on Technology-Facilitated Gender-Based Violence: M.Ş.D v Romania' (*Strasbourg Observers*, 22 April 2025) <https://strasbourgobservers.com/2025/04/22/strasbourgs-consolidation-on-technology-facilitated-gender-based-violence-m-s-d-v-romania/> accessed 3 May 2025.

¹⁸⁹ Catherine Van de Heyning, 'Strasbourg's Consolidation on Technology-Facilitated Gender-Based Violence: M.Ş.D v Romania' (*Strasbourg Observers*, 22 April 2025) <https://strasbourgobservers.com/2025/04/22/strasbourgs-consolidation-on-technology-facilitated-gender-based-violence-m-s-d-v-romania/> accessed 3 May 2025.

In 2003, the CoE adopted the Additional Protocol to the Convention on Cybercrime (Budapest Convention)¹⁹⁰, becoming one of the first binding international instruments to tackle online hate speech. Extending the scope of the Convention, the Protocol required States Parties to criminalise the dissemination of racist and xenophobic material through computer systems (Article 1), explicitly covering expressions based on race, colour, descent, national or ethnic origin, and religion (Article 2). However, as evident from its title and scope, and reflection of the historical context of its adoption, the Protocol excluded a wide range of protected characteristic – among them ‘gender’.¹⁹¹ Despite leaving a significant gap in European-level legal protection against ICT-facilitated forms of violence, the Protocol marked a foundational moment in the international legal response to online hate. It remains the only instrument specifically designed to address online hateful conduct, setting a precedent for the criminalisation of certain forms of harmful expression in the digital sphere.¹⁹²

The Istanbul Convention (2011)

The CoE’s Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) is widely regarded as a benchmark in the fight against gender-based violence. However, it does not explicitly reference hate speech based on gender, nor is this issue addressed in its Explanatory Report or in the preparatory work of the Ad Hoc Committee (CAHVIO) responsible for drafting the Convention.¹⁹³

Despite this, the Convention’s broad definition of violence against women – as a form of human rights violation and discrimination – encompasses gender-based acts likely to cause physical, sexual, psychological, or economic harm.¹⁹⁴ Within this framework, it is

¹⁹⁰ 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* (opened for signature 28 January 2003, entered into force 1 March 2006) ETS No 189.

¹⁹¹Natalie Alkiviadou, ‘Hate Speech on Social Media Networks: Towards a Regulatory Framework?’ (2019) 28(1) *Information & Communications Technology Law* 19, 23.

¹⁹² Ibid 28.

¹⁹³ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2338.

¹⁹⁴ Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention) (opened for signature 11 May 2011, entered into force 1 August 2014) CETS No 210, Art 3(a).

possible to argue that sexist hate speech falls under the scope of ‘psychological harm’ and, thus, ‘violence against women’.¹⁹⁵

Drawing inspiration from Article 5 CEDAW, Article 12 requires state parties to eliminate prejudices, customs, traditions, and all other practices grounded in the notion of women’s inferiority or in stereotyped gender roles.¹⁹⁶ Accordingly, this provision seems to oblige states not only to combat sexist hate speech, but also to address sexist speech more broadly, even when it does not meet the threshold of hate speech.¹⁹⁷ The obligations set out in Article 12 are further elaborated in the subsequent provisions of Chapter III of the Convention.

Although Articles 33 to 40 of the Istanbul Convention require the criminalization of various forms of violence against women, only certain types of sexist hate speech may be captured under these provisions. Articles 34 (stalking) and 40 (harassment)¹⁹⁸ have been identified as particularly relevant when addressing hate speech.¹⁹⁹ However, these two articles are insufficient to address the full scope of sexist hate speech. Article 34 applies only when such speech constitutes part of a repeated and threatening pattern that induces fear in the targeted individual. Article 40 is limited to cases where the ‘unwanted verbal conduct’ is explicitly sexual in nature. Yet sexist hate speech warrants legal recognition even when it does not exhibit a systematic pattern, and it need not be sexualised to cause harm²⁰⁰ - as previously discussed, it may involve non-sexual, but still gendered, insults or expressions of gender-based hostility. What is more, sexist hate speech often occurs in public settings, in contrast to stalking and sexual harassment, which typically take place in more private or personal contexts.²⁰¹

This gap in legal coverage could potentially be addressed by Article 33, which requires state parties to criminalise ‘psychological violence’, defined as intentional behaviour aimed at seriously harming an individual’s psychological well-being through coercion or

¹⁹⁵ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2338.

¹⁹⁶ Ibid.

¹⁹⁷ Ibid.

¹⁹⁸ Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021) para 39.

¹⁹⁹ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2339.

²⁰⁰ Ibid.

²⁰¹ Ibid.

threats.²⁰² Nevertheless, as clarified by the treaty drafters, Article 33 applies to ‘an abusive pattern of behaviour occurring over time’, rather than isolated incidents.²⁰³

Another important question is how these provisions apply in an online context. In this respect, GREVIO has emphasised the importance of media self-regulation in curbing the spread of sexist and hateful content.²⁰⁴ Although an explicit reference to the online dimension of violence against women can only be found in Article 17 (Participation of the Private Sector and the Media), which calls for the active involvement of ICT sectors, including digital platforms, GREVIO has interpreted Articles 33, 34 and 40 as encompassing forms of online and technology-facilitated abuse.²⁰⁵

Two conclusions can be drawn from this. Firstly, the Istanbul Convention lacks a provision that unequivocally requires states to criminalise or otherwise sanction all forms of sexist hate speech. However, this specialised treaty's value as a reference for integrating a gender perspective into the work of international human rights bodies, such as the ECHR, should not be underestimated, particularly with regard to hate speech.²⁰⁶

Secondly, although an explicit reference to the online dimension of gender-based violence within the text of the Convention would have been preferable, GREVIO's interpretation in GR No. 1 is nonetheless significant. It demonstrates that the interpretation of the Convention is capable of evolving to address emerging forms of gender-based violence in the digital sphere, thereby ensuring that these increasingly prevalent manifestations remain within the Convention's scope.

3.1.1. The Specificity of Political Contexts

When a female politician is subjected to sexist hate speech or disinformation campaigns, such attacks – whether isolated or repeated – can deter her from engaging actively in

²⁰² Ibid.

²⁰³ Ibid; Council of Europe, *Explanatory Report to the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence* (2011) para 181.

²⁰⁴ Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021), para 57(g)(h).

²⁰⁵ Ibid para 36.

²⁰⁶ Sękowska-Kozłowska, Baranowska and Grabias (n 98) 2339.

public discourse and debate, including on social media platforms. Beyond violating the rights to non-discrimination, human dignity, and potentially the right to private life,²⁰⁷ these attacks may also infringe on her fundamental political rights – notably, the right to participate in public affairs, including the right to stand for election.

At the same time, regarding alleged hate speech, when targeting a public figure, international standards dictate that it is not enough for an expression to simply be deemed insulting in order to justify imposing penalties. Although public figures, including those in the highest political offices, like heads of state and government, are entitled to the protections provided by instruments such as the ECHR, they must also accept that, as part of their role, they may legitimately be subject to heightened levels of public criticism.²⁰⁸

Therefore, at the CoE level, in cases regarding sexist attacks against female politicians – none of which have yet been adjudicated by the ECtHR –, it becomes crucial to distinguish legitimate political critique – fully protected under Article 10(1) ECHR as key in a democratic society – from misogynistic speech, which inherently lacks any public interest. This entails identifying when a comment goes beyond being merely 'offensive, shocking, or disturbing'²⁰⁹ and instead advocates, promotes or incites to violence or discrimination.

Although defamation and hate speech are distinct concepts – defamation primarily addressing individual reputation under Article 8 ECHR, and hate speech involving both individual harm and broader societal concerns such as democracy, public order, and the protection of vulnerable groups²¹⁰ - the Court's reasoning in defamation cases remains relevant to the analysis of hate speech. This is because the Court often applies similar standards in cases involving politicians, particularly when assessing the necessity of a

²⁰⁷ The ECtHR has recognised that the unique characteristics of the internet can be taken into account when assessing the severity of an attack on a person's reputation, in order to determine whether it falls within the scope of Article 8 (*Arnarson v Iceland* App no 58781/13, ECtHR, 13 September 2017, para 37).

²⁰⁸ *Lingens v Austria* App no 9815/82 (ECtHR, 8 July 1986) para 42; *Nikula v Finland* App no 31611/96 (ECtHR, 21 June 2002) para 48; *Palomo Sánchez and Others v Spain* [GC] Apps nos 28955/06, 28957/06, 28959/06 and 28964/06 (12 September 2011) para 71.

²⁰⁹ Additionally, in a case concerning a politician, the Court has established that to even constitute defamation expression must be more than just 'vulgar abuse', that is, offensive language commonly found online (*Tamiz v United Kingdom* App no 3877/14, ECtHR, 19 September 2017, para 81).

²¹⁰ Federica Casarosa, *Freedom of Expression and Countering Hate Speech: Handbook on Techniques of Judicial Interaction in the Application of the EU Charter* (European University Institute 2021) 23.

restriction under Article 10(2). For example, concepts such as the distinction between value judgments and statements of fact, the importance of contextual analysis, and the heightened protection afforded to free speech in political debate – typically developed in defamation cases – have been employed to support the Court’s reasoning in hate speech cases as well.²¹¹

Concerning the distinction between judgements of value and statements of fact,²¹² the Court has clarified that expressing opinions, or value judgments, in an online context cannot in itself constitute defamation.²¹³ Nevertheless, the Court has also noted that even when a statement constitutes a value judgment, it must nonetheless be grounded in an adequate factual basis; otherwise, it risks being deemed disproportionate or unjustified.²¹⁴ This is relevant in excluding hate speech from protection in the sense that misogynistic expressions could never be grounded in ‘an adequate factual basis’, since their very nature is rooted in social prejudices and fed on malicious intent.

Following this logic, in order to identify hate speech in such cases, it is possible to draw from the Court’s case-law a three-level base approach:

- (i) Identify if a statement is ‘only’ offensive, shocking or disturbing, or if it incites to violence, hatred or hostility; in order to do so,
- (ii) Assess whether the speech contributed to a matter of public interest, thus falling within the scope of freedom of expression necessary in a democratic and pluralistic society, or whether it amounted to a deliberately provocative statement lacking any contribution to public debate.

²¹¹ See, for example, *Erbakan v Turkey* App no 59405/00 (ECtHR, 6 October 2006) para 65.

²¹² In *Lingens v Austria* (1986), the Court observed that ‘[t]he existence of facts can be demonstrated, whereas the truth of value judgements is not susceptible of proof’ (*Lingens v Austria* App no 9815/82, ECtHR, 8 July 1986, para 46). See also *Morice v France* [GC] App no 29369/10 (ECtHR, 23 April 2015) para 126, *Dalban v Romania* [GC] App no 28114/95 (ECtHR, 28 September 1999) para 49 and *Oberschlick v Austria* (no 1) App no 11662/85 (ECtHR, 23 May 1991) para 63.

²¹³ *Pedersen and Baadsgaard v Denmark* [GC] App no 49017/99 (ECtHR, 17 December 2004) para 76; *De Haes and Gijssels v Belgium* App no 19983/92 (ECtHR, 24 February 1997) para 42; *Oberschlick v Austria* (no 2) App no 20834/92 (ECtHR, 1 July 1997), para 33; *Lindon, Otchakovsky-Laurens and July v France* [GC] Apps nos 21279/02 and 36448/02 (ECtHR, 22 October 2007) para 55.

²¹⁴ *Ibid.*

- (iii) Understand the context, the intention of the speaker, and what impact did the statement likely have.²¹⁵

Regarding point (i), the distinction between sexist speech that, although oppressive and damaging, does not amount to hate speech and sexist hate speech becomes crucial.²¹⁶ In this context, particular attention should be given to the fact that incitement to hatred does not necessarily entail calls for violence or other criminal acts, but may consist of speech that ‘merely’ insults, ridicules, or slanders specific groups, as held by the ECtHR.²¹⁷

Regarding point (ii), as noted above, the democratic necessity for public officials to tolerate a higher level of criticism should be considered. Still, Article 8 of the Convention, which safeguards the right to respect for private life, entails the protection of an individual's reputation, including that of public figures such as politicians. Thus, this protection remains applicable even when politicians are not acting in their private capacity. What distinguishes these contexts is the necessity of carefully balancing the protection of reputation with the fundamental interest in preserving open and uninhibited political debate.²¹⁸ Therefore, a value judgment may lose the protection of Article 10 once it qualifies as hate speech, given the need to safeguard individual reputation and, more broadly, human dignity, the right to non-discrimination, and democratic values under Article 10(2).

Lastly, point (iii) calls for a contextual, case-by-case analysis, which should, moreover, guide the overall assessment. In cases of sexist hate speech, the repeated use of gendered slurs (as discussed in section 2.2.1) to target a female politician serves to sexualise, objectify, and dehumanise, revealing a clear intent to harm her public image. This not only poses a risk of psychological harm but may also discourage her continued participation in public life, including her expression on social media platforms.²¹⁹

²¹⁵ The emphasis on the 'likely' impact is crucial in that, as discussed in relation to the definition of VAWIP in Chapter II, it shifts the focus toward the speaker's motive and intention rather than relying solely on the actual effects of the statement – an approach that risked being overly reductive.

²¹⁶ Richardson-Self (n 84) 263.

²¹⁷ *Vejdeland and Others v Sweden* App no 1813/07 (ECtHR, 9 February 2012) para 55.

²¹⁸ See *Lingens v Austria* App no 9815/82 (ECtHR, 8 July 1986) para 42; *Nadtoka v Russia* App no 38010/05 (ECtHR, 31 May 2016) para 42.

²¹⁹ To exemplify, a study examining the social media activity of politically active women on X (formerly Twitter) before and after they were subjected to online attacks revealed compelling evidence that such abuse

Furthermore, the ECtHR has determined that the greater the impact of speech, the greater its potential to disturb public order.²²⁰ This reasoning can be applied to the context of social media by considering factors such as the platform's reach, the number of views, reactions, comments, or shares a post it receives.

In conclusion, when assessing sexist attacks against female politicians, it is crucial to distinguish between legitimate political critique and misogynistic hate speech, recognising that the latter – inherently devoid of public interest and rooted in discriminatory intent – not only harms personal reputation and dignity but also undermines broader democratic values, and must therefore be subject to stricter scrutiny under a context-sensitive, multi-level analysis grounded in evolving ECtHR jurisprudence.

3.2. The EU

At the EU level, efforts to address hate speech are reflected across primary, secondary, and soft law instruments.

In an effort to comprehensively conceptualise the phenomenon, the European Institute for Gender Equality (EIGE) defines online gender-based hate speech as content disseminated via ICT that: (a) expresses hatred toward women and/or girls on the basis of their gender, or due to the intersection of gender with other characteristics such as race, age, disability, sexual orientation, ethnicity, nationality, religion, or profession; and/or (b) incites, promotes, justifies, or spreads hatred grounded in gender or intersecting factors.²²¹

The Charter of Fundamental Rights of the European Union (CFREU) is particularly relevant in the context of treaty law. Although freedom of expression is strongly

discourages women from maintaining their engagement on the platform (National Democratic Institute for International Affairs, *Tweets that Chill: Analyzing Online Violence Against Women in Politics*, 2019, 21).

²²⁰ *Karataş v Turkey* App no 23168/94 (ECtHR, 8 July 1999) para 52; Matthias C. Kettermann and Wolfgang Benedek, 'Freedom of Expression Online' in Mart Susi (ed), *Human Rights, Digital Society and the Law: A Research Companion* (Routledge 2019) 125.

²²¹ European Institute For Gender Equality, *Cyber Violence against Women and Girls: Key Terms and Concepts* (EIGE 2022) 6.

safeguarded under its Article 11(1), which mirrors Article 10(1) ECHR, it is still subject to limitations that may arise from the need to balance it with other fundamental rights.

To this effect, Article 52(1) CFREU makes it mandatory to ensure that such restrictions are provided by law and Article 52(3) stipulates that the rights within the Charter must be interpreted with the same meaning and scope as the corresponding rights in the ECHR. Additionally, Article 6(2) of the Treaty on European Union (TEU) affirms the EU's commitment to accede to the ECHR. Although this commitment has not yet materialised and remains controversial, it suggests that the EU's fundamental standards on hate speech are intended to closely align with those of the CoE.²²²

The limitations imposed by the right to equality and non-discrimination to the exercise of freedom of expression derive foundationally from Article 2 TEU, which provides that 'The European Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities'. Article 10 of the Treaty on the Functioning of the European Union (TFEU) further states that, when developing and carrying out its policies and actions, the EU must strive to eliminate discrimination based on sex, race or ethnic background, religion or beliefs, disability, age, or sexual orientation. Additionally, Article 19 TFEU grants the EU the authority to take suitable measures to fight against discrimination on these same grounds.²²³ The CFREU reinforces and develops such guarantees in its Article 21 (non-discrimination) – which widens the TFEU's scope of protected characteristics to encompass social origin, genetic features, language, political or any other opinion, membership of a national minority, property and birth – and Article 23, which addresses specifically the principle of equality between men and women.

²²² Consolidated Version of the Treaty on European Union [2012] OJ C326/15, Art 6(2); Eva Nave and Lottie Lane, 'Countering Online Hate Speech: How Does Human Rights Due Diligence Impact Terms of Service' (2023) 51 *Computer Law & Security Review* 1, 4.

²²³ Beatrix Immenkamp, Ionel Zamfir and David de Groot, *Hate Speech and Hate Crime: Time to Act?* (European Parliamentary Research Service, 2024) 6.

3.2.1. Towards expanding the scope of gender-based cyberviolence to include sexist hate speech

Although the EU's accession to the CoE's Istanbul Convention on 1 October 2023 represented a notable development in combating gender-based violence, its legal effect remains confined to areas within EU competence.²²⁴ Since several Member States have yet to ratify the Convention,²²⁵ the need for a cohesive and legally binding instrument at the EU level became, thus, both apparent and essential to ensure consistent protection and the advancement of gender equality across the Union.

Inspired by the principles of equality and non-discrimination enshrined in the TEU and the CFREU, the European Parliament adopted a resolution in 2021 formally recognising gender-based violence as a 'new area of crime' under Article 83(1) TFEU.²²⁶ Importantly, this resolution enabled the EU to take legislative measures to establish minimum rules concerning the definition and nature of sanctions for gender-based criminal offences of a cross-border dimension. In a related resolution the same year the Parliament urged the Commission and Member States to broaden the definition of gender-based cyberviolence to encompass sexist hate speech, and to explicitly recognise misogyny as a ground for hate speech.²²⁷ These efforts materialised in the adoption of the recent Directive (EU) 2024/1385 on Combating Violence Against Women and Domestic Violence, which represented a significant milestone in paving the way for harmonised legislation in the field of gender-based violence, including hate speech.²²⁸

²²⁴ Court of Justice of the European Union, *Opinion 1/19* (Grand Chamber, 6 October 2021), ECLI:EU:C:2021:799, [273].

²²⁵ As of 2025, the Member States that have not ratified the Istanbul Convention are: Bulgaria, Czechia, Hungary, Latvia, Lithuania and Slovakia. In July 2020, the Polish government declared its intention to withdraw from the Convention, although no formal withdrawal has taken place to date (European Parliament, 'EU Accession to the Istanbul Convention', *Legislative Train Schedule*, 6 June 2023) <https://www.europarl.europa.eu/legislative-train/theme-a-new-push-for-european-democracy/file-eu-accession-to-the-istanbul-convention> accessed 10 June 2025).

²²⁶ European Parliament, 'Resolution of 16 September 2021 with recommendations to the Commission on identifying gender-based violence as a new area of crime listed in Article 83(1) TFEU' [2021] OJ C117/12, Annex, Art 1.

²²⁷ European Parliament, 'Resolution of 14 December 2021 with recommendations to the Commission on combating gender-based violence: cyberviolence' [2021] OJ C 251.

²²⁸ European Parliament, 'Resolution of 16 September 2021 on the EU Strategy for Gender Equality' [2021] OJ C 456, para 17.

Article 8 of the Directive is particularly relevant, as it explicitly mandates Member States to criminalise incitement to violence or hatred against a group of persons, or a member of such a group, defined by reference to gender, by publicly disseminating material containing such incitement online.²²⁹

Consistent with what was already outlined in section 2.2.1., the Directive underscores the importance of identifying a 'biased motivation', emphasising that such bias may not always be explicit in the acts themselves but can be inferred from their broader content and contextual circumstances.²³⁰

However, the second paragraph of Article 8 grants Member States significant discretion in transposing its provisions into national law, allowing them to penalize 'only conduct which is either carried out in a manner likely to disturb public order or which is threatening, abusive or insulting'.²³¹ The final part of Recital 27 clarifies that where national law imposes additional conditions, these must not compromise the effectiveness of the provision criminalising cyber incitement to violence or hatred.²³² Still, the flexibility afforded by the provision, while intended to accommodate national legal systems, opens the door to different interpretations of such broad concepts. In my view, a more specific approach – such as that adopted in the CoE's 2022 Recommendation on Combating Hate Speech, which clearly identifies the forms of conduct that should be criminalised²³³ – would have been preferable in ensuring stronger protection and greater legal harmonisation across jurisdictions.

Furthermore, from a broader perspective, the Directive falls short of fully adopting a human rights-based approach to gender-based violence – a foundational principle of the Istanbul Convention.²³⁴ This shortcoming is particularly evident when comparing Article

²²⁹ Immenkamp, Zamfir and De Groot (n 223) 7.

²³⁰ Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Recital 25.

²³¹ Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Art 8(2).

²³² Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Recital 27.

²³³ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, para 11.

²³⁴ Ceren Kasim, 'Advancing Gender Equality: The EU's Landmark Directive 2024/1385 on Violence Against Women' (2024) *SSRN* 1, 5.

2(a) of the Directive with Article 3(a) of the Istanbul Convention, where the latter explicitly frames gender-based violence as a violation of human rights.²³⁵ By failing to adopt this perspective, the Directive missed a critical opportunity to align more closely with the Istanbul Convention and to set a transformative precedent in addressing gender-based and domestic violence across the EU.²³⁶

Nevertheless, this framework encouraged significant reforms in Member States' legal systems and a wider cultural shift in how gender inequality and violence is recognised and addressed throughout the EU, serving as a catalyst for further progress on the issue of gendered hate speech.²³⁷ It will be important to observe the subsequent developments, as Member States are obligated to transpose the Directive into national legislation by 14 June 2027.²³⁸

3.2.2. The European Commission's Proposal to Extend the list of 'EU crimes' to include Hate Speech (2021)

In her 2020 State of the European Union Address, Commission President Ursula Von der Leyen announced that the Commission would propose 'to extend the list of EU crimes to all forms of hate crime and hate speech, whether because of race, religion, gender or sexuality'.²³⁹ Indeed, a crucial development, which preceded Directive (EU) 2024/1385, was reflected in the 2021 European Commission Communication 'A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime', aimed at securing a decision from the Council of the European Union on expanding the scope of Article 83(1) TFEU.²⁴⁰ A proposal for such Council decision is annexed to the Communication. Should such a decision be adopted, the Commission would then have competence to propose substantive legislation harmonising the definition of and penalties

²³⁵ Ibid.

²³⁶ Ibid.

²³⁷ Ibid.

²³⁸ Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36], Art 49. The Directive will be further analysed in section 3.2.1.

²³⁹ European Commission, 'A More Inclusive and Protective Europe: Extending the List of EU Crimes to Hate Speech and Hate Crime' COM(2021) 777 final, 9 December 2021.

²⁴⁰ Ibid.

for hate speech and hate crime based on gender/sex. In addition to complementing Directive (EU) 2024/1385, this proposal forms part of a wider range of initiatives by the Commission, including the EU Gender Equality Strategy (2020-2025). The Strategy explicitly recognises the need to include sex-based hate speech and hate crime in the list of so-called 'Eurocrimes'.²⁴¹ It recognises these acts as specific forms of gender-based violence, emphasising that 'too many people still violate the principle of gender equality by engaging in sexist hate speech and obstructing action against gender-based violence and gender stereotypes'.²⁴²

The Council's Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law²⁴³ is currently the only EU measure that standardises how certain types of hate speech and hate crimes are defined and punished in EU Member States jurisdictions.²⁴⁴ As a result of transposing the Framework Decision into national law, hate speech should now be criminalised in all Member States on grounds of race, colour, religion, descent, national or ethnic origin.²⁴⁵ However, the Framework Decision does not extend protection to other important and widely recognised personal characteristics such as sex, gender, sexual orientation, age, disability, or socio-economic status, thereby leaving up entirely to the Member States to explicitly criminalise hate speech on such grounds.²⁴⁶ This leads to significant variation in legal protection across the EU, including in what concerns gender-based hate speech.

In fact, hate speech frameworks differ in both scope and severity, with notable variations in the very definition of hate speech, the enumerated protected grounds, the criteria required to establish hate speech offences, and the practical enforcement and application of these provisions at the national level.²⁴⁷ Up to 2021, only 17 Member States were

²⁴¹ Ibid 86.

²⁴² European Commission, 'A Union of Equality: Gender Equality Strategy 2020–2025' COM(2020) 152 final, 5 March 2020, 2.

²⁴³ Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L328/55.

²⁴⁴ Sara de Vido and Lorena Sosa, *Criminalisation of Gender-Based Violence against Women in European States, Including ICT-Facilitated Violence* (European Commission, Directorate-General for Justice and Consumers 2021) 149.

²⁴⁵ Ibid.

²⁴⁶ Ibid.

²⁴⁷ Peršak (n 90) 100.

reported to criminalise hate speech on grounds of sex/gender²⁴⁸ – among these, 14 specifically referred to the ground of sex/gender, whereas 10 referred (alternatively or in addition) to the ground of gender identity and two referred to the ground of ‘sex characteristics’.²⁴⁹

When it comes specifically to hate speech through ICT means, a 2022 study by EIGE found that gender is only referenced explicitly as one of the grounds of hate speech in 10 EU countries: Estonia, Greece, Spain, Latvia, Lithuania, Hungary, Malta, Austria, Portugal and Slovenia.²⁵⁰ Moreover, research indicates that only a limited number of states have enacted legislation explicitly addressing the online dimension of hate speech as either a constituent element of the offence or an aggravating factor, through references to the internet, media, or information and communication technologies.²⁵¹

While some Member States seem to be looking to criminalise gendered hate speech and crime, following its perceived increase both globally and in Europe, others remain hesitant, citing concerns over excessive restriction of freedom of expression.²⁵² As a reflection of the absence of harmonised legal recognition, although prosecution for criminal hate speech is generally *ex officio*, national experts report a notably low incidence of judicial proceedings related to hate speech in general, and in particular, those concerning sexist hate speech.²⁵³ This reflects a deeper, systemic issue: a lack of awareness among public authorities regarding the legal parameters for identification and an insufficient understanding of gender-specific dynamics in detecting this form of violence against women.²⁵⁴

²⁴⁸ This means that 10 countries do not explicitly criminalise hate speech on the grounds of sex or gender. These countries are: Bulgaria, Czechia, Denmark, Finland, Germany, Ireland, Italy, Poland, Romania and Slovakia. Nevertheless, in the case of Germany the ‘indirect’ incorporation of gender can be found in Article 130 of the Criminal Code where ‘other segment of the population’ is mentioned; in the case of Italy, while debates have been light up around the proposed Zan Bill, repeatedly blocked by the Senate, developments in this regard are yet to be achieved (De Vido and Sosa, n 244, 154, 156).

²⁴⁹ European Commission, ‘A More Inclusive and Protective Europe: Extending the List of EU Crimes to Hate Speech and Hate Crime’ COM(2021) 777 final, 9 December 2021, 13.

²⁵⁰ European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls* (EIGE 2022) 53.

²⁵¹ De Vido and Sosa (n 244) 157.

²⁵² European Commission, ‘Roadmap: Hate Speech & Hate Crime – Inclusion on List of EU Crimes’ (*Better Regulation Portal*, 2020) https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12872-Hate-speech-&-hate-crime-inclusion-on-list-of-EU-crimes_en accessed 28 April 2025, 3.

²⁵³ De Vido and Sosa (n 244) 157.

²⁵⁴ *Ibid.*

The Proposal

While the existing list of ‘EU crimes’ set out in Article 83(1) TFEU is exhaustive, it may be extended in accordance with the procedure outlined in its third subparagraph.²⁵⁵ As established in Article 83(1) TFEU, EU crimes refer to ‘(i) areas of (ii) particularly serious crimes with (iii) a cross-border dimension resulting from the nature or impact of such offences or from a special need to combat them on a common basis’. The present list of such areas of crime largely correspond to the areas that were subject of third-pillar legislation (Cooperation in the Fields of Justice and Home Affairs), including ‘terrorism, trafficking in human beings and sexual exploitation of women and children, illicit drug trafficking, illicit arms trafficking, money laundering, corruption, counterfeiting of means of payment, computer crime and organised crime’.

In this regard, it is relevant to briefly ascertain how hate speech meets the criteria set out in this provision in order to be considered an EU crime.

(i) Area of crime

Hate speech and hate crime have been considered one ‘area of crime’ – instead of two separate ‘areas’ – by the European Commission, as they share an intrinsic feature which is the motivation rooted in hatred.²⁵⁶

(ii) Particularly seriousness of the harm

Concerning the required ‘particularly seriousness of the harm’, although large-scale research on the effects of hate crime and hate speech is still limited, existing studies reveal that these acts cause significant and long-lasting harm.²⁵⁷ What is more, these actions affect not only the direct victim but also the wider community that shares the targeted trait, highlighting their serious nature.²⁵⁸ In fact, the Commission has referred to such

²⁵⁵ ‘On the basis of developments in crime, the Council may adopt a decision identifying other areas of crime that meet the criteria specified in this paragraph. It shall act unanimously after obtaining the consent of the European Parliament.’ (Consolidated Version of the Treaty on the Functioning of the European Union [2012] OJ C326/47, Art 83(1), para 3).

²⁵⁶ European Commission, ‘Commission Work Programme 2021: A Union of Vitality in a World of Fragility’ COM(2020) 690 final, 19 October 2020, 7; European Commission, ‘A More Inclusive and Protective Europe: Extending the List of EU Crimes to Hate Speech and Hate Crime’ COM(2021) 777 final, 9 December 2021, 7.

²⁵⁷ Peršak (n 90) 111.

²⁵⁸ Ibid.

harms in three main categories: (i) harm to the individual victim, (ii) harm to the victim's community, and (iii) harm to broader society.²⁵⁹

(iii) Cross border dimension

Under Article 83(1) of the TFEU, a crime may be considered cross-border in nature due to the way it is committed, the effects it produces across countries, or because there is a particular need for a coordinated response at the EU level 'on a common basis'.²⁶⁰ In the case of online hate speech, the most prevalent form of hate speech, the transnational dimension in nature and impact is evident considering that cyberspace knows no borders.²⁶¹ Nevertheless, the need to combat it 'on a common basis' is an equally strong argument. A rise in hate crime and hate speech might also deter people from exercising their right to free movement – one of the core freedoms under EU law – by making them less willing to relocate to EU Member States that are perceived as hostile for their lack of protection through criminal law.²⁶²

That said, cautions need to be taken not to overextend the reach of criminalisation measures. The emphasis should stay mainly on the first group (the individual directly targeted) – and possibly the second group (the individuals' community) – of victims who suffer direct and immediate harm.²⁶³ This approach should also be consistent with the CoE's standards on the gradation of sanction severity.²⁶⁴ As Nina Peršak logically argues, it is the immediate harm that should form the basis for justifying criminalisation and, in turn, any EU action under Article 83 TFEU, as only such harm should warrant a punitive response from the state or the EU.²⁶⁵

The success of this initiative in guaranteeing the identical hate speech criminalisation across all Member States is crucial in light of the principle of subsidiarity, which requires

²⁵⁹ Ibid.

²⁶⁰ Ibid 113.

²⁶¹ Ibid 114.

²⁶² Ibid.

²⁶³ Ibid 113.

²⁶⁴ Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16, paras 3-4, 7-9.

²⁶⁵ Ibid 113.

careful consideration within the context of criminal law, and reflects a strong commitment to core EU values enshrined in Article 2 TEU.²⁶⁶

In December 2023, the European Commission and High Representative called on the Council to swiftly add hate speech and hate crime to the list of EU crimes.²⁶⁷ The European Parliament echoed this in a January 2024 resolution.²⁶⁸ Despite these efforts, as of July 2025, the Council has yet to reach the required unanimity to adopt the proposal – a necessary step before a Directive can be adopted. This institutional deadlock leaves gender-based hate speech largely unregulated at the EU level, resulting in fragmented and inconsistent national responses.²⁶⁹ This is in stark contrast to the wider trend of greater gender sensitivity within EU frameworks.

Legal foundations such as Article 2 TEU, Article 21 CFREU, and Article 19 TFEU – which explicitly prohibit discrimination on the basis of sex and guide the Union’s functional competences – reflect the EU’s core competence in addressing gender-based discrimination and, therefore, should serve as primary reference points for identifying relevant categories of crime under Article 83(1) TFEU. As hate speech (and hate crime) represent some of the most serious and harmful manifestations of bias and discrimination they should be regarded as natural extensions of the EU’s anti-discrimination principles, applied within the realm of criminal law.²⁷⁰ For these reasons, the grounds protected under anti-discrimination law and those covered by hate speech legislation ought to be fully aligned. Hence, following Nina Peršak’s reasoning, I argue that the exclusion of ‘gender’ or ‘sex’ from the scope of EU hate speech legislation lacks logical justification.²⁷¹

In this context, the European Commission’s new initiative represents a significant and necessary advancement toward addressing this gap and strengthening the protection of

²⁶⁶ Peršak (n 90) 101.

²⁶⁷ European Commission and High Representative of the Union for Foreign Affairs and Security Policy, ‘Joint Communication to the European Parliament and the Council – No Place for Hate: A Europe United Against Hatred’ JOIN(2023) 51 final, 6 December 2023; Immenkamp, Zamfir and De Groot (n 223) 10.

²⁶⁸ European Parliament, ‘Resolution of 18 January 2024 on Extending the List of EU Crimes to Hate Speech and Hate Crime (2023/2068(INI))’ (2024).

²⁶⁹ European Commission, ‘Proposal for a Directive on Combating Violence Against Women and Domestic Violence’ COM(2022) 105 final, Explanatory Memorandum, Objectives of the Proposal, para 4.

²⁷⁰ Peršak (n 90) 91.

²⁷¹ Peršak (n 90) 86.

the rights of individuals, particularly those that are targets of gender-based discrimination.²⁷²

The lack of a coherent EU framework is especially detrimental to women in politics, who remain frequent targets of online and offline gender-based attacks that weaken democratic participation and erode legal commitments to equality in public life.²⁷³

3.3. Gendered Disinformation: at the Crossroads of Protected and Unprotected Speech

In section 2.2.2 of the previous Chapter, I argued that gendered disinformation targeting female politicians should be recognised as a hybrid phenomenon, both in its element of information disorder and its element of gender-based violence. However, addressing the issue in its full scope within a specific legal framework poses further challenges. The absence of a clear legal qualification of gendered disinformation, contributes to legal ambiguity in determining the appropriate framework under which it should be classified. This results in a dual risk: addressing the phenomenon only as an information disorder may fail to capture its gendered nature, while framing it exclusively as gender-based violence may fail to acknowledge its broader societal and democratic implications.

International free speech protections do not permit restrictions on expression solely on the basis of falsity. This principle is grounded in a well-documented concern that laws against so-called ‘false news’ have frequently been misused by authoritarian governments to suppress dissent and silence criticism, rather than to genuinely counter disinformation harmful to the public.²⁷⁴ Thus, prohibiting speech merely for being false remains incompatible with international standards on freedom of expression.

Nevertheless, gendered disinformation, because it involves coordinated efforts and malicious intent, can have tangible and harmful effects on the rights and dignity of those

²⁷² Ibid 96.

²⁷³ Zamfir (n 13) 1-3.

²⁷⁴ Pramukh Nanjundaswamy Vasist, Debashis Chatterjee and Satish Krishnan, ‘The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-Country Configural Narrative’ (2023) *Information Systems Frontiers* 1, 14.

targeted. Therefore, in the perspective of addressing individual cases, it may meet the threshold for a legitimate restriction under Article 10(2) ECHR, both of which permit limitations on expression to protect the ‘rights of others’. In this context, gendered disinformation may, like hate speech, justify legal intervention when it undermines the dignity, psychological integrity and freedom of expression of those it targets.

In the recent *Avagyan v. Russia* (2025) the ECtHR addressed online disinformation for the first time, concerning the context of the Covid-19 pandemic. In confronting Article 10 ECHR with the interest of protecting public health, the Court assessed whether the national courts provided relevant and sufficient reasons for charging the applicant with the offence of ‘disseminating untrue information’.²⁷⁵ The Court established that, for a restriction on freedom of expression to be proportionate, the offence must meet a high threshold, requiring that the information in question be: (a) known to be untrue; (b) socially significant; (c) presented under the guise of reliable reporting; and (d) likely to pose specific risks to public health, safety, or infrastructure.²⁷⁶ This stringent standard reflected the Court’s caution in endorsing legal measures against disinformation, underscoring the previously noted risk of enabling state control over information.²⁷⁷ The ECtHR’s reasoning highlights the delicate balance between combatting harmful disinformation practices and ensuring that restrictions do not lead to excessive limitations on freedom of expression.²⁷⁸ It also leaves open the possibility for future refinement or reinforcement of the established standards in forthcoming cases.²⁷⁹

Currently, disinformation is addressed through a patchwork of policy instruments and soft law mechanisms.²⁸⁰ In the digital realm, emerging legally binding instruments such as the EU’s Digital Services Act (DSA) have begun to pay particular attention to the risks it poses, especially where systemic harms are involved – for example, public health, public

²⁷⁵ Babette De Naeyer, ‘Avagyan v. Russia: Strasbourg’s Appetizer on Online Disinformation’ (*Strasbourg Observers*, 27 June 2025) <https://strasbourgobservers.com/2025/06/27/avagyan-v-russia-strasbourgs-appetizer-on-online-disinformation/> accessed 27 June 2025.

²⁷⁶ *Ibid*; *Avagyan v Russia* App no 36911/20 (ECtHR, 29 April 2025) para 32.

²⁷⁷ A concurring opinion by three judges underscored this tension, warning against the danger of allowing states to become ‘arbiters of “truth”’ (De Naeyer, n 275).

²⁷⁸ De Naeyer (n 275).

²⁷⁹ *Ibid*.

²⁸⁰ At the EU level, see for example the Communication on Tackling Online Disinformation (2018), the Action Plan Against Disinformation (2018) and the Code of Practice on Disinformation (2018, updated 2022).

security or the integrity of democratic processes. However, when gender is a central element, the complexity of such cases is heightened, making legal classification more challenging and increasing the risk of underrecognition – particularly where the normalization of sexism obscures the harmful nature of the conduct.

Depending on the specificities of each case and the national jurisdiction, gendered disinformation may fall under several different frameworks, ranging from anti-discrimination to digital platforms' due diligence obligations.²⁸¹ In some cases – particularly those involving reputational harm or privacy violations – civil remedies, such as defamation suits, may offer the most appropriate and proportionate recourse.²⁸² Furthermore, although it is important to highlight that gendered disinformation does not automatically constitute hate speech,²⁸³ it can be argued that it may cross into this domain in egregious scenarios – particularly when it incites hostility, discrimination, or violence on the basis of gender.²⁸⁴ Such cases are characterised not merely by the dissemination of falsehoods with malicious intent and some degree of coordination, but by their use of dehumanising, sexualised, or aggressive language intended to humiliate, intimidate or incite to hatred. A notable example is the disinformation campaign against former Ukrainian MP Svitlana Zalishchuk, where a fabricated tweet containing a false statement was circulated alongside doctored nude images.²⁸⁵ This attack weaponised sexual shaming and ridicule to discredit her, reflecting a clear intent to incite gender-based contempt and social exclusion – hallmarks of hate speech. In such circumstances, it may be both appropriate and necessary to apply the more stringent legal frameworks specifically designed to address hate speech.

In any case, given the frequently contested boundary between hate speech or defamation and protected expression, any legal response must proceed with caution.²⁸⁶ It ought to be

²⁸¹ Wilfore (n 118) 126.

²⁸² Bardall (n 71) 119.

²⁸³ Ibid 112.

²⁸⁴ Irene Khan, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (7 August 2023) UN Doc A/78/288, para 24.

²⁸⁵ Jankowicz (n 134).

²⁸⁶ European Economic and Social Committee, 'Freedom of Expression Is Not a Licence to Engage in Hate Speech' (2022) <https://www.eesc.europa.eu/en/news-media/news/freedom-expression-not-licence-engage-hate-speech> accessed 12 June 2025.

firmly rooted in contextual analysis and proportionality, ensuring that restrictions do not overreach, while still adequately addressing the seriousness of such harms.

In this light, gendered disinformation – particularly when directed at women in public life – represents a serious concern not only for the individuals affected but also for the health of democratic discourse more broadly. By distorting public debate, deterring female participation, and eroding the foundational principle of equal representation, it raises substantive challenges for democratic societies.²⁸⁷

Legal and policy responses must therefore reflect the complex and multifaceted nature of this phenomenon, carefully balancing the protection of freedom of expression with the imperative to prevent gender-based harm and threats to democratic processes. In this context, the broader societal and political implications of gendered disinformation should be considered as aggravating factors even when addressing the phenomenon through anti-discrimination or defamation laws. Simultaneously, the gender perspective should be advanced in general frameworks aimed at combating information disorders. Such a dual approach allows for a proportionate and context-sensitive response – one that effectively addresses both the individual harms inflicted and the broader risks posed to the integrity of democratic foundations.

Chapter IV: Regulating Gendered Hate Speech and Disinformation on Social Media

Acknowledging the dual nature of the Internet as both an enabler of expression and a vector for abuse, the ECtHR has held that imposing obligations on digital platforms, including social media, to remove unlawful content upon notification is consistent with the ECHR.²⁸⁸ In cases involving hate speech, the Court has gone further, imposing

²⁸⁷ Wilfore (n 118) 131.

²⁸⁸ This precedent was first established in *Delfi AS v Estonia*, which concerned liability for unlawful comments posted on internet news portals, with the Court clarifying that its reasoning was not intended to encompass the context of social media platforms (*Delfi AS v Estonia*, GC, App no 64569/09, ECtHR, 16

heightened duties of care that require platforms to go beyond notice-and-take-down systems and act proactively in detecting and removing such content.²⁸⁹

In the EU context this is no longer a matter of legal uncertainty, as the DSA²⁹⁰ now clearly sets out the duty to moderate content.²⁹¹

Considering the demonstrated impact of social media content on public discourse, electoral processes²⁹² and offline violence, content moderation has been acknowledged as a matter of public interest rather than a purely private enterprise. Consensus around the insufficiency of self-regulation and the need for independent oversight and external intervention was articulated by European MPs in four arguments: (i) ‘What is illegal offline has to be illegal online’; (ii) the EU is ‘taking back control’ from big and US based enterprises;²⁹³ (iii) the EU is ‘protecting small businesses, consumers, and our citizens against big tech’; and (iv) it is developing ‘a golden standard and rulebook beyond the EU’.²⁹⁴

Nevertheless, substantial challenges endure in addressing gender-based hate speech and disinformation within the contours of an evolving regulatory landscape. The dual capacity

June 2015, paras 110, 115-116). However, similar reasoning and principles were subsequently extended to social media platforms (e.g. *Sanchez v France* App no 45581/15, ECtHR, 15 May 2023, paras 86, 140, 190).

²⁸⁹ For example, in *Sanchez v France* the Court held that a minimum level of moderation or prior filtering by the host platform or account holder, aimed at identifying and removing ‘clearly unlawful content’ within a reasonable time, may be required, even in the absence of a prior complaint by the injured party (*Sanchez v France* App no 45581/15, ECtHR, 15 May 2023, paras 86, 140, 190); Matthias C. Kettermann and Wolfgang Benedek, ‘Freedom of Expression Online’ in Mart Susi (ed), *Human Rights, Digital Society and the Law: A Research Companion* (Routledge 2019) 104.

²⁹⁰ Entered into force on 16 November 2022 and became fully applicable to all online platforms on 17 February 2024.

²⁹¹ Therese Enarsson, ‘Navigating Hate Speech and Content Moderation under the DSA: Insights from ECtHR Case Law’ (2024) 33 *Information & Communications Technology Law* 384, 392.

²⁹² The European Parliament and the CoE have recognised the dangers posed by the weaponization of hate and disinformation to undermine democratic processes (European Parliament, ‘Resolution of 25 November 2020 on Strengthening Media Freedom: The Protection of Journalists in Europe, Hate Speech, Disinformation and the Role of Platforms’ [2020] OJ C 425/51, para W).

²⁹³ Importantly, the DSA applies to providers of intermediary services irrespective of their place of establishment or residence (Recital 7).

²⁹⁴ Gabi Schlag, ‘European Union’s Regulating of Social Media: A Discourse Analysis of the Digital Services Act’ (2023) 11(3) *Politics and Governance* 168, 169. Niklas Eder, ‘Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation’ (2024) 25 *German Law Journal* 1197, 1204;

of AI – to both amplify these harms and bolster the precision of their detection and mitigation – introduces an additional layer of complexity.

4.1. The EU Digital Services Act

Building on the foundations of the E-Commerce Directive,²⁹⁵ the DSA introduces a harmonised EU framework aimed at addressing systemic digital risks, protecting fundamental rights, and enhancing the accountability of online platforms. It introduces tailored due diligence responsibilities based on a platform’s role, size, and impact, and obliges intermediary service providers to integrate fundamental rights considerations into their terms of service (Article 14(4)), mandating that they conduct broader assessments of how their systems may negatively impact such rights at a structural level (Article 34).²⁹⁶

The complementary Codes of Conduct developed are particularly relevant for addressing both hate speech and disinformation. By incorporating, in January 2025, the updated version of the Code of Conduct on Countering Illegal Hate Speech Online (originally adopted in 2016), the DSA reinforces platform obligations to detect, assess, and remove illegal hate speech, while promoting greater transparency and accountability in content moderation.²⁹⁷ This integration means that participating platforms can now be held accountable under the DSA for commitments made through the Code. In contrast, the DSA stops short of formally incorporating the strengthened 2022 Code of Practice on Disinformation developed under the European Democracy Action Plan.²⁹⁸ Nonetheless, its relevance is acknowledged and Recital 104 states that ‘the refusal without proper explanations by a provider of an online platform [...] of the Commission’s invitation to

²⁹⁵ Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1.

²⁹⁶ Ibid 1206.

²⁹⁷ European Commission, ‘Code of Conduct on Countering Illegal Hate Speech Online’ (*Digital Strategy*, 2025) <https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online> accessed 16 June 2025.

²⁹⁸ European Parliament, ‘European Democracy Action Plan’ (*Legislative Train Schedule*) <https://www.europarl.europa.eu/legislative-train/theme-a-new-push-for-european-democracy/file-european-democracy-action-plan> accessed 27 June 2025.

participate in the application of such a code of conduct could be taken into account, where relevant, when determining whether the online platform [...] has infringed the obligations laid down by this Regulation’.²⁹⁹ In Article 45, the DSA further incentivizes adherence to developed Codes of Conduct by linking it to enhanced perceptions of transparency, risk mitigation, and enforcement – particularly for signatories classified as Very Large Online Platforms (VLOPs), which are subject to annual independent audits.

4.1.1. Defining ‘Illegal Content’

Article 3(h) DSA adopts a broad definition of ‘illegal content’:

“‘illegal content’” means any information that, in itself or in relation to an activity [...] is not in compliance with Union law or the law of any Member-State which is in compliance with Union law, irrespective of the precise subject matter or nature of that law’³⁰⁰

This reliance on national legal definitions introduces fragmentation in classification, since Member States diverge considerably in their approaches to the scope and substance of what constitutes illegal hate speech – e.g. an offensive comment shared on X might be prosecuted as a criminal offence in Germany but acceptable and unpunished in France or Portugal.³⁰¹

Recital 12 of the DSA offers some clarification in this regard. It expands the concept of illegal content to include ‘information relating to such illegal content’ and provides illustrative examples such as ‘hate speech’ and ‘unlawful discriminatory content’. However, despite these explicit references, the DSA does not define hate speech, nor does it establish a conceptual or legal framework for its identification and regulation. An attempt to fill this definitional vacuum might look to the Code of Conduct on Countering

²⁹⁹ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 104.

³⁰⁰ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Art 3(h).

³⁰¹ Gabi Schlag, ‘European Union’s Regulating of Social Media: A Discourse Analysis of the Digital Services Act’ (2023) 11(3) *Politics and Governance* 168, 170.

Illegal Hate Speech Online, as it was formally integrated into the DSA framework.³⁰² However, the Code’s hate speech definition relies on Article 1(1) of Framework Decision 2008/913/JHA, which frames hate speech exclusively in terms of racism and xenophobia.³⁰³

Hence, the same harmonization gap previously discussed prevails, particularly considering that only a subset of Member States explicitly recognises gender as a protected characteristic in hate speech frameworks.³⁰⁴ As outlined in the previous Chapter, although Article 8 of Directive (EU) 2024/1385 obliges Member States to criminalise serious forms of sexist hate speech, the broad discretion it affords Member States in aligning their national laws will likely result in uneven implementation, undermining the DSA’s capacity to effectively address the full spectrum of gender-based abuse online.³⁰⁵ This challenge is compounded by the fluid and context-dependent nature of content on social media. As a result, misogynistic material may be classified as ‘unlawful discriminatory content’ but not consistently recognised as hate speech. This broader categorisation, however, risks obscuring the specific nature of misogyny and downplaying its distinct social, psychological, and structural harms.

This gap becomes particularly acute when the abuse is directed at female politicians. In such cases, perpetrators might exploit the robust protections granted to freedom of expression in political contexts, claiming that even overtly sexist remarks or incitements to violence constitute legitimate contributions to public discourse. This strategic weaponisation of free speech norms poses a serious barrier to accountability, complicating efforts by both platforms and regulators to respond effectively.³⁰⁶

³⁰² Enarsson (n 291) 392.

³⁰³ Ibid 400.

³⁰⁴ European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls* (EIGE 2022) 53.

³⁰⁵ For example, the DSA’s notice-and-action mechanisms may prove ineffective in legal systems where sexist speech is not explicitly criminalised, or where enforcement is weak or ambiguous (Enarsson, n 291, 399).

³⁰⁶ Reuters, ‘Meta’s Instagram Failed to Curtail Hate Speech against Women Politicians, Report Says’ (14 August 2024) <https://www.reuters.com/technology/metas-instagram-failed-curtail-hate-speech-against-women-politicians-report-says-2024-08-14/> accessed 22 June 2025.

One way to mitigate legal fragmentation would be to recognise gender-based hate speech as an EU crime, as proposed by the European Commission in 2021.³⁰⁷ Such a development would help standardise definitions and enforcement obligations, ensuring that serious cases of gendered hate speech are clearly captured under the illegal content provisions of the DSA. It would also compel platforms to address the gender dimension of harmful content in a legally binding and enforceable manner, thereby enabling a more robust regulatory response to online misogyny targeting female politicians in the EU.

Regarding disinformation, although the DSA references the phenomenon in several recitals,³⁰⁸ it does not acknowledge that it might amount to illegal content³⁰⁹ nor explicitly mention it in the regulation's binding provisions – aside from its inclusion in risk assessment obligations. Nevertheless, building on previous analysis, it is crucial to emphasise that disinformation, when gendered and particularly when directed at female politicians, can and should be understood as a form of gender-based violence. Classifying such disinformation as illegal content, however, presents additional challenges. These arise from the lack of a clear or consistent definition of the phenomenon and from the DSA's reliance on national legal frameworks: in some Member States, the same content might be treated under anti-discrimination frameworks or as a minor case of defamation, while in others it could constitute criminal hate speech.

4.1.2. Due Diligence Obligations

Recital 46 clarifies that the due diligence obligations under the DSA are designed to protect key public policy objectives, particularly the safety and trust of users – especially those who are vulnerable to hate speech, sexual harassment, and discrimination. This framing highlights the importance of recognising and addressing gender-based risks as an integral part of interpreting and implementing the DSA.

³⁰⁷ Discussed in section 3.2.2. of this thesis.

³⁰⁸ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recitals 2, 9, 69, 83, 84, 88, 95, 104, 106, 108.

³⁰⁹ See, for example, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 84.

Crucially, the DSA establishes a heightened duty of care for VLOPs,³¹⁰ recognising their broader societal impact. According to Article 34 (complemented by Recital 80), VLOPs must exercise due diligence in identifying, analysing, and assessing content that is likely to constitute a ‘systemic risk’ (Article 34). These systemic risks are divided into four categories – succinctly: (a) the dissemination of illegal content; and ‘any actual or foreseeable negative effects’ (b) to the exercise of fundamental rights’ as enshrined in the CFREU³¹¹, (c) ‘on civic discourse and electoral processes, and public security’ or (d) ‘in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person’s physical and mental well-being’.³¹²

Gender-based hate speech might fit into all the categories outlined above, since it: (i) constitutes illegal content, as explicitly clarified in the interpretative effort of Recital 12; (ii) it can have ‘actual or foreseeable negative effects’ to the exercise of freedom of expression of the female politician targeted (Article 11 CFREU) as well as her right to non-discrimination compared to her male counterparts (Article 21 CFREU) and the fundamental right to human dignity (Article 1 CFREU); (iii) it can harm civic discourse and electoral processes by distorting the public image of a legitimate political candidate based on a personal attribute – her sex – thereby undemocratically compromising her public support; (iv) as outlined in previous Chapters, it constitutes VAWIP, and therefore gender-based violence, in that it is capable of producing gender-motivated psychological harm to the individual and, more broadly, a chilling effect on women’s political participation and freedom of expression.

While the DSA explicitly recognises gender-based violence and hate speech as priority concerns, its treatment of disinformation – particularly when gendered – is notably less clear. Recital 83 mentions coordinated disinformation campaigns as a relevant risk, yet

³¹⁰ Those that exceed an average of 45 million active recipients monthly, equivalent to 10% of the EU population, calculated over a period of six months. (Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 76). E.g. Facebook (Meta), Instagram (Meta) or X (X Corp.).

³¹¹ See also Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 81.

³¹² See also Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 83.

Article 34 does not include a direct reference to the issue. This suggests that such campaigns would be implicitly included under Article 34(c), which addresses negative effects on civic discourse and electoral processes. When disinformation is gendered and intersects with gender-based violence, it may also fall within the scope of Article 34(b) and (d) for the same reasons as gender-based hate speech.

As such, digital platforms must account for both gendered hate speech and gendered disinformation in their mandatory risk assessments.

Nevertheless, the absence of explicit legal grounding for disinformation within the systemic risk framework suggests a reluctance by the EU legislator to regulate this issue through binding obligations. Instead, the approach often relies on soft law instruments, intended as ‘guiding principles’, which ultimately afford platforms considerable discretion in defining and enforcing their Terms of Service. This flexibility risks leaving gendered disinformation insufficiently addressed – not only because effective regulation of the ‘disinformation’ aspect remains underdeveloped, but also because its gendered dimension is still largely overlooked, especially in political contexts where such dynamics are easily obscured or dismissed.

Beyond risk identification, Article 35 DSA mandates that platforms implement specific mitigation measures to address the identified risks.³¹³ These measures must be reasonable, proportionate, and effective, and tailored to the systemic risks in question.³¹⁴ This includes revising content moderation practices to improve the speed and accuracy of responses to certain types of illegal content.³¹⁵ In severe cases – such as instances of illegal hate speech or cyberviolence,³¹⁶ an area where women are disproportionately affected – platforms are obliged to promptly remove or restrict access to the harmful content. This reflects a concrete duty to take preventive action, not merely to assess risk.³¹⁷

³¹³ Claudia Morini, ‘Countering Online Sexist Hate Speech in the European Legal Context: Between Present Commitment and Future Challenges’ (2024) 104 *QIL* 17, 29.

³¹⁴ *Ibid.*

³¹⁵ *Ibid.*

³¹⁶ *Ibid.*

³¹⁷ *Ibid.*

Transparency obligations further support due diligence. Articles 15, 24, and 42 DSA (complemented by Recital 65), require platforms to publish periodic transparency reports.

With the integration of the Code of Conduct on Countering Illegal Hate Speech into the DSA, transparency and reporting standards have been strengthened. Signatories are now encouraged to enhance their efforts in four key areas: (i) Reporting on the outcomes of mitigation measures; (ii) Providing additional data on hate speech, including insights into the influence of recommender systems and the organic or algorithmic reach of harmful content prior to its removal; (iii) Presenting disaggregated, country-level data on hate speech, categorised by factors such as gender identity, sexual orientation, race, and ethnicity; (iv) Ensuring meaningful multi-stakeholder follow-up and cooperation.³¹⁸

Notwithstanding these developments, the lack of clear definitions and enforcement guidelines for identifying and moderating gender-based hate speech and disinformation continues to be a critical shortcoming. As a result, platform compliance often centres on procedural formalities, such as demonstrating that risk management systems are in place, rather than ensuring alignment with substantive, rights-based standards.³¹⁹ This procedural focus undermines effective enforcement. Platforms may appear compliant while neglecting to embed gender-sensitive approaches into their terms of service or content moderation practices. In this regard, EIGE's contribution to the public consultation on the DSA raises important concerns about reporting templates, highlighting the need to explicitly include incitement to gender-based hatred and violence – an element not currently mandated within the DSA's reporting framework.³²⁰ EIGE further recommended that reporting metrics capture the prevalence and specific forms of online gender-based hate speech, to ensure the gender dimension is properly integrated into systemic risk assessments.³²¹

³¹⁸ European Commission, *Code of Conduct+ on Countering Illegal Hate Speech Online: Integration into the Digital Services Act Framework* (20 January 2025) <https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online-plus> accessed 19 June 2025.

³¹⁹ Eva Nave and Lottie Lane, 'Countering Online Hate Speech: How Does Human Rights Due Diligence Impact Terms of Service' (2023) 51 *Computer Law & Security Review* 1, 2.

³²⁰ European Institute for Gender Equality, *EIGE's contribution to the open public consultation 'Digital Services Act – transparency reports (detailed rules and templates)'* (EIGE 2024).

³²¹ *Ibid.*

From a wider perspective, the societal harms posed by such content, extending beyond harm to individual users, must be more clearly acknowledged and integrated into risk assessment practices in order to proportionately address the severity of its impact.³²²

4.1.3. Content Removal Obligations

In accordance with Article 6(1)(b) and Article 9(1) DSA – supported by Recital 22 – online platforms are obligated to take prompt and effective measures once they are notified of illegal content hosted on their services. Complementing this, the Code of Conduct on Countering Illegal Hate Speech, which has been signed by 12 companies to date – including seven VLOPs: Facebook, Instagram, LinkedIn, Snapchat, TikTok, X, and YouTube³²³ – requires signatories to assess the majority of valid notifications concerning illegal hate speech within 24 hours and, where appropriate, to remove or block access to the flagged content.³²⁴

Given the inherent challenges in identifying hate speech, user-generated alerts – particularly those submitted by ‘trusted flaggers’ – play a vital role in enabling timely interventions. This aligns with the reporting and redress mechanisms outlined in Articles 16 (notice and action mechanisms) and 22 (trusted flaggers) of the DSA.

However, a recurring issue persists: the lack of a clear, harmonised definition of ‘hate speech’, combined with the Code’s limited focus on racist and xenophobic expressions, creates ambiguity regarding the exact scope of content subject to moderation and the corresponding removal obligations.³²⁵

³²² Niklas Eder, ‘Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation’ (2024) 25 *German Law Journal* 1197, 1205.

³²³ European Commission, ‘Digital Services Act – Codes of Conduct’ <https://digital-strategy.ec.europa.eu/en/policies/dsa-codes-conduct> accessed 19 June 2025.

³²⁴ European Commission, ‘Code of Conduct on Countering Illegal Hate Speech Online’ (*Digital Strategy*, 2025) <https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online> accessed 19 June 2025. Article 52 provides a quantitative threshold for penalties for infringements that Member States should observe (Article 52(2)(3)), determining that these penalties should be ‘effective, proportionate and dissuasive’ (Article 52(2)).

³²⁵ Enarsson (n 291) 400.

Like hate speech, disinformation lacks a universally accepted legal definition within the regulatory framework, leading to challenges in delineating platform responsibilities. Even when it is not considered illegal per se within a particular national jurisdiction, disinformation can have significant detrimental effects on democratic processes and public trust, particularly when weaponised against female politicians based on gender, and still require some form of regulation. With over 40 signatories,³²⁶ the 2022 revised version of Code of Practice on Disinformation (originally adopted in 2018) contains 44 voluntary commitments and 128 specific measures that platforms should adopt to tackle false or misleading content that can harm public discourse.³²⁷ The Code requires companies to clearly label political ads, remove fake accounts, help users make informed decisions, strengthen monitoring, and reduce ad revenue for accounts and websites that spread false information.³²⁸ However, although most platforms' terms of service prohibit threats of violence, hate speech, manipulated media and deepfakes, fake accounts, and coordinated inauthentic behaviour, evidence from digital forensic investigations reveals that gendered disinformation, despite breaching these rules, continues to flourish on their platforms.³²⁹

4.1.4. The Obligation to Remove 'Identical' or 'Equivalent' Content

Article 8 of the DSA reinforces the fundamental principle established in Article 15 of the E-Commerce Directive that prohibits Member States from imposing a general obligation on platforms to proactively monitor user activity or to engage in active fact-finding for potentially illegal content. However, the DSA framework introduces a more dynamic regulatory approach by explicitly allowing intermediary service providers to issue targeted, binding orders for the removal of specifically identified illegal content on their own initiative (Article 7 DSA). This novelty introduced by Article 7 DSA, in contrast to

³²⁶European Commission, 'Digital Services Act – Codes of Conduct' <https://digital-strategy.ec.europa.eu/en/policies/dsa-codes-conduct> accessed 19 June 2025.

³²⁷European Commission, 'Code of Practice on Disinformation' (*Digital Strategy*) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> accessed 21 June 2025.

³²⁸ Ibid.

³²⁹ Wilfore (n 118) 133.

the E-Commerce Directive, clarifies that engaging in voluntary diligence does not compromise the liability exemptions of online platforms for content they did not originate, modify, or influence, as granted under Articles 4, 5, and 6 DSA.

Notably, the European Court of Justice (ECJ) case *Glawischnig-Piesczek v. Facebook* (2019), which concerned defamatory content targeting a female politician, played a pivotal role in shaping this development. Eva Glawischnig-Piesczek, a former Austrian MP and leader of the Green Party, was targeted in a Facebook post that commented on a linked news article featuring her image, accusing her of being part of a ‘fascist party’, a ‘corrupt oaf’, and a ‘traitor to her people’ (*Volksverräterin*) – a term historically associated with Nazi rhetoric.³³⁰ Because the post originated from a user operating under a false name, Glawischnig-Piesczek was unable to take legal action against the individual directly. Instead, she submitted a takedown request to Facebook, citing the offensive and defamatory nature of the content.³³¹ Facebook declined to remove the post, prompting her to seek legal recourse through the Austrian courts.³³² The courts ultimately ruled in her favour, finding the comments to constitute unlawful defamation under Austrian law.

However, when the case reached the Austrian Supreme Court, it raised several questions concerning the interpretation of the E-Commerce Directive, which were referred to the ECJ for a preliminary ruling. One key issue was whether requiring Facebook to remove not only the original post but also all identical or equivalent content posted by other users would violate the prohibition against imposing general monitoring obligations on intermediaries, as set out in Article 15 of the Directive. In this regard, the ECJ held that requiring platforms to remove content that is ‘identical’ or ‘equivalent’ to the content primarily identified as illegal does not violate Article 15. Consequently, these injunctions may extend to content that ‘remains essentially unchanged’ or contains substantially the same elements.³³³ In accordance with the principle that cyberspace ‘knows no borders’,

³³⁰ Clara Rauchegger and Aleksandra Kuczerawy, ‘Injunctions to Remove Illegal Online Content under the eCommerce Directive: *Glawischnig-Piesczek*’ (2020) 57(5) *Common Market Law Review* 1495, 1496.

³³¹ *Ibid.*

³³² *Ibid.*

³³³ Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Ltd* [2019] ECLI:EU:C:2019:821, paras 41, 45, 46 and 53. The judgment extended beyond the Advocate General’s opinion, which held that only the removal of ‘identical content’ – excluding ‘equivalent content’ uploaded by other users – would comply with Article 15 E-Commerce Directive (Opinion of A.G. Szpunar in Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, EU:C:2019:458, para 73).

the ECJ further recognised that this EU law obligation can be enforced beyond the territorial scope of the EU.³³⁴

Beyond broadening the interpretation of platform obligations, this case was a pioneer in recognising the severity of systematic and widespread attacks on female politicians. It emphasised the urgent need to enhance efforts to address such attacks effectively, particularly given the viral nature of sharing and resharing on social media, in order to ensure adequate protection for the victims.

Another groundbreaking judgment involving a female politician – *Renate Künast v. Facebook* (2021)³³⁵ – extended this reasoning at the national level. Renate Künast, a member of the German Bundestag representing the Green Party and a long-standing target of online gendered hate and sexually explicit speech,³³⁶ was this time targeted by a widely shared and reposted defamatory meme on Facebook. The meme falsely attributed to her the statement: ‘Integration starts with you as a German learning Turkish!’³³⁷ Although Künast had already denied making the statement back in 2015, when she requested Facebook to remove the defamatory meme along with identical or similar misquotations, Facebook only blocked the original meme.³³⁸ However, the German courts reaffirmed the *Glawischnig-Piesczek* ECJ precedent and additionally ruled that platforms may be required to use automated detection tools to identify and remove similar infringing

³³⁴ ‘Article 15(1), must be interpreted as meaning that it does not preclude a court of a Member State from: [...] ordering a host provider to remove information covered by the injunction or to block access to that information worldwide within the framework of the relevant international law’ (Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Ltd* [2019] ECLI:EU:C:2019:821, para 53).

³³⁵ *Renate Künast v Facebook* (District Court Frankfurt, 2-03 O 188/21, 2021).

³³⁶ A long legal dispute over 22 comments of sexist and violent nature posted on Facebook about a distorted statement made by Künast in 1986 came to an end in December 2021, when the Federal Constitutional Court overturned the Berlin Courts’ decisions, finding that her general personality rights and the users’ freedom of expression had not been properly balanced (the Courts had found that only 12 of the 22 comments were potential crimes), emphasising that public servants deserve effective protection from hate speech (DW, ‘Facebook Abuse: Top German Court Rules for Lawmaker’, DW, 26 January 2022, <https://www.dw.com/en/facebook-abuse-top-german-court-rules-for-lawmaker/a-60634812> accessed 21 June 2025; Global Freedom of Expression, ‘The Case of Renate Kunast’, *Columbia University*, 15 October 2019, <https://globalfreedomofexpression.columbia.edu/cases/the-case-of-renate-kunast/> accessed 21 June 2025).

³³⁷ Global Freedom of Expression, ‘Kunast v Facebook’ (*Columbia University*, 7 January 2022) <https://globalfreedomofexpression.columbia.edu/cases/kunast-v-facebook/> accessed 24 June 2025.

³³⁸ *Ibid.*

content once they have been notified of its illegality and provided with a list of unlawful expressions to target.³³⁹

The outcomes of these cases seem to have significantly influenced the formulation and interpretation of Article 7 DSA, which formalises the power of national authorities to issue targeted, binding orders to remove specific illegal content or prevent access to it, including across the EU. Together, Articles 7 and 8 DSA preserve the principle that platforms should not be subject to an obligation of enforcing general monitoring while codifying an important legal expectation that, once having identified illegal content, they must act not only against the original content but also against its foreseeable replications.

These developments are particularly relevant to ensure that, once gendered hate speech or disinformation – the latter being clearly issue in the case of Kunäst – is recognised as ‘illegal’ under national jurisdiction, enforcement is effective. This is crucial to combat the widespread wave of gendered hate speech and disinformation targeting female politicians on social media platforms, which consistently damages their reputations, careers, and psychological well-being.

4.2. The Role of Artificial Intelligence

The advancement of AI technologies has introduced both opportunities and complexities in addressing gender-based abuse on social media, including the spread of hate speech and disinformation. On the one hand, machine learning models – such as AI-driven moderation systems and content filters – often perpetuate algorithmic bias. Combined with the rise of generative AI, including the proliferation of bots³⁴⁰ in political discourse, these technologies have significantly contributed to the growing backlash against women’s public presence and expression on social media, particularly that of female politicians. On the other hand, they have demonstrated significant potential to enhance

³³⁹ Ibid.

³⁴⁰ According to the Oxford Internet Institute, ‘bots’ are automated accounts designed to mimic human activity on social media. They are frequently used to artificially amplify certain narratives or suppress opposing voices, for example by spreading hate speech or manipulated information (Samantha Bradshaw and Philip N Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation* (Computational Propaganda Research Project, Oxford Internet Institute 2019) 1, 11).

the precision and effectiveness of moderating illegal content disseminated across social media platforms.

Risks to amplifying gendered abuse

Although structural and cultural sexism is central to understanding political violence against women, social norms alone do not fully account for the rapid rise of online attacks targeting female politicians.³⁴¹

Social media platforms' algorithms increasingly rely on machine learning models designed to prioritize content likely to drive high user engagement. In doing so, companies often privilege profit maximization over human rights due diligence, frequently neglecting to verify the accuracy of content, assess its potential harm, or consider its broader societal impact.³⁴² As a result, algorithmic bias³⁴³ - shaped by user behaviour and embedded in these models – can lead AI systems to replicate and amplify the dissemination of misogynistic content, particularly targeting public figures such as female politicians.³⁴⁴

Furthermore, the proliferation of generative AI significantly heightens pre-existing risks associated with cyberviolence against women and girls.³⁴⁵ These technologies facilitate new forms of abuse such as automated generation of hateful messages and harassment

³⁴¹ Wilfore (n 118) 132.

³⁴² Ibid 133.

³⁴³ 'Algorithmic bias' refers to the unintentional perpetuation and amplification of societal biases embedded in machine learning processes or the training data used in their development (Marie Lamensch, 'Generative AI tools are perpetuating harmful gender stereotypes' (*Center for International Governance Innovation*, 14 June 2023) <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/> accessed 22 June 2025).

³⁴⁴ European Institute for Gender Equality, *Combating Cyberviolence Against Women and Girls: Developing an EU Measurement Framework* (EIGE 2025) 44-45; Rebecca Heilweil, 'Why Algorithms Can Be Racist and Sexist' (*Vox*, 18 February 2020) <https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency> accessed 22 June 2025; Wilfore (n 118) 132; Nicola Henry and Asher Flynn, 'Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support' (2019) 25 *Violence Against Women* 1932, 1949-51; Ouyang and others (n 135) 250.

³⁴⁵ European Institute for Gender Equality, *Combating Cyberviolence Against Women and Girls: Developing an EU Measurement Framework* (EIGE 2025) 44-45.

and the creation of synthetic content nearly indistinguishable from that produced by humans.³⁴⁶

This poses risks of manipulation and unprecedented spread of hatred and discrimination on social media platforms.³⁴⁷ One such harm involves the use of deepfake technology to produce hyper-realistic, non-consensual sexual imagery or audio, severely affecting victims' reputations and psychological well-being.³⁴⁸ Additionally, AI-powered chatbots and virtual assistants can reflect and reinforce gender biases, sometimes producing inappropriate or sexually explicit content that results in manipulation of public discourse, particularly in political contexts.³⁴⁹

The final paragraph of Article 34 DSA mandates that risk assessments conducted by online platforms must include an analysis of whether these risks are 'influenced by intentional manipulation of their service, including by inauthentic use or automated exploitation of the service, as well as the amplification and potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions'. In this context, watermarking techniques intended to identify whether something was created by a generative model have proved to be useful tools, yet they face notable limitations, such as poor compatibility across different social media platforms.³⁵⁰

³⁴⁶ Ibid.

³⁴⁷ Parliamentary Assembly of the Council of Europe, *Resolution 2590 (2025) on regulating content moderation on social media to safeguard freedom of expression* (adopted 30 January 2025) Res 2590 (2015) para 12.

³⁴⁸ Equality Now, 'The rise of deepfake image-based sexual abuse necessitates urgent and comprehensive responses from technological innovation, legal reform, and societal awareness' (Equality Now, 2024) https://equalitynow.org/news_and_insights/the-rise-of-deepfake-image-based-sexual-abuse-necessitates-urgent-and-comprehensive-responses-from-technological-innovation-legal-reform-and-societal-awareness/ accessed 21 April 2025.

³⁴⁹ M Heikkilä, 'The Viral AI Avatar App Lensa Undressed Me – Without My Consent' *MIT Technology Review* (12 December 2019) <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/> accessed 21 April 2025; European Institute for Gender Equality, *Combating Cyberviolence Against Women and Girls: Developing an EU Measurement Framework* (EIGE 2025) 44-45. Furthermore, in Recital 104, the DSA recognises that disinformation can manifest through 'the use of bots or fake accounts for the creation of intentionally inaccurate or misleading information, sometimes with a purpose of obtaining economic gain' (Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1, Recital 104).

³⁵⁰ Parliamentary Assembly of the Council of Europe, *Resolution 2590 (2025) on regulating content moderation on social media to safeguard freedom of expression* (adopted 30 January 2025) Res 2590 (2025) para 12.

The recent AI Act marks a significant milestone in the EU's efforts to regulate artificial intelligence through binding legislation. Adopting a risk-based framework, the Act categorizes AI systems into four tiers of risk and assigns corresponding obligations primarily to providers, deployers, manufacturers, and other actors in the AI value chain. These categories include: (i) Unacceptable risk systems, which are prohibited outright (e.g., social scoring or manipulative AI); (ii) High-risk systems, which are subject to strict compliance requirements, including risk management, data governance, human oversight, and conformity assessments; (iii) Limited-risk systems, which must meet transparency obligations (e.g., users must be informed when interacting with AI); and (iv) Minimal-risk systems, which are largely unregulated under the Act.³⁵¹

Importantly, the AI Act³⁵² imposes robust obligations for AI systems that may affect fundamental rights,³⁵³ including requirements for bias mitigation, transparency, and human oversight – essential safeguards to reduce the amplification of gender-based hate speech and disinformation.³⁵⁴ It also mandates the labelling of AI-generated content, such as deepfakes (Article 50(4)), a growing concern given their use in targeting female public figures. By establishing these legal guardrails, the AI Act complements the DSA, jointly addressing the systemic risks posed by AI to democratic participation and gender equality online.

Despite this highly positive development, and while secondary legislation is not adopted, significant enforcement challenges persist. Monitoring compliance across diverse platforms and jurisdictions is complex, especially when powerful generative models are developed and deployed by actors outside the EU.³⁵⁵ Technical limitations – such as ineffective watermarking, limited interoperability, and evolving evasion tactics –

³⁵¹ European Commission, 'Regulatory Framework for Artificial Intelligence' (*Digital Strategy*) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 2 July 2025.

³⁵² Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L168/1.

³⁵³ Bird & Bird, *European Union Artificial Intelligence Act: A Guide* (7 April 2025) 3.

³⁵⁴ *Ibid.*

³⁵⁵ According to Recital 22 and Article 2 AI Act, non-EU providers whose AI outputs are used in the EU are also subject to its rules.

undermine transparency measures.³⁵⁶ Furthermore, the reliance on self-assessments by AI providers and fragmented national enforcement may dilute accountability, making it difficult to respond swiftly and effectively to the harms disproportionately affecting women in digital political spaces.

Automated Content Moderation

While emergent regulations like the DSA highlight a strengthened role of the State in setting standards for moderating illegal content on social media, platforms – and even users – still play a primary role in regulating speech, often more directly than the courts.³⁵⁷ In this context, the current framework gives them broad discretion to design and adjust moderation systems, including how they balance automation and human oversight.³⁵⁸ Recognising this reality, the DSA requires transparency regarding the use of automated systems in content moderation (e.g. Articles 15, 16, and 17).

In 2019, CoE’s PACE had already stated that social media platforms ought to use algorithms designed to enhance the variety of sources, subjects, and perspectives presented to users, ensure the reliability of the information provided, and, in doing so, help minimise the dangers of ‘filter bubbles’ and ‘echo chambers’.³⁵⁹ Indeed, automation enables quicker responses by using upload filters and actively enforcing ‘notice and take-down’ requirements. Furthermore, both the Code of Practice on Disinformation and the accompanying Communication advocate the integration of AI systems as a tool to effectively combat disinformation.³⁶⁰ Empirical research suggests that limiting the frequency of content sharing is an effective way to curb the spread of disinformation

³⁵⁶ Justyna Lisinska and Daniel Castro, ‘Why AI-Generated Content Labeling Mandates Fall Short’ (Center for Data Innovation, 2024) 1, 1-4.

³⁵⁷ Enarsson (n 291) 385-386.

³⁵⁸ Ibid.

³⁵⁹ Parliamentary Assembly of the Council of Europe, *Resolution 2281 (2019) Social Media: Social Threats or Threats to Human Rights?* (adopted 12 April 2019) Res 2281 (2019) para 9.7.; Parliamentary Assembly of the Council of Europe, *Resolution 2590 (2025) on regulating content moderation on social media to safeguard freedom of expression* (adopted 30 January 2025) Res 2590 (2025), para 16.

³⁶⁰ Viktoria Mazur and Archil Chochia, ‘Definition and Regulation as an Effective Measure to Fight Fake News in the European Union’ (2022) 9(1) *European Studies – the Review of European Law, Economics and Politics* 15, 28.

while safeguarding freedom of expression;³⁶¹ the use of AI could further enhance this approach.

However, relying solely on AI for content moderation presents significant risks. AI systems may misinterpret context, leading to over-censorship or inconsistent and biased flagging of content, raising concerns about their ability to distinguish between harmful and harmless content, false and accurate information.³⁶² The highly personal (and gendered) dimension of hate speech and targeted disinformation may further complicate fact-checking processes.³⁶³

This concern is echoed in a report by Cambridge Consultants, which suggests a more balanced approach to content moderation.³⁶⁴ The report proposes a two-step model in which AI systems conduct an initial pre-moderation assessment before content is published. Content deemed to be free of obvious harmful material is then passed on for human review.³⁶⁵ Human moderators can subsequently adjust the AI's decision-making processes through machine learning, enabling the system to better identify inconsistencies in future content.³⁶⁶ This iterative process promotes a collaborative model in which AI assists in filtering and fact-checking, while human oversight ensures accuracy and contextual understanding.³⁶⁷ In this process it is crucial that a gender-sensitive perspective is incorporated.

AI has also proven to be a useful tool in user-reported assistance. Regarding hate speech, AI-driven chatbots impersonating other users have been used to encourage users be more accountable for their own posts and to report harmful content created by others.³⁶⁸ These same tools could be applied to disinformation, confronting users with trustworthy information. This is perhaps an underexplored strategy that holds great potential to counter the dissemination of hate speech and disinformation online.

³⁶¹ Ibid.

³⁶² Ibid.

³⁶³ Wilfore (n 118) 137.

³⁶⁴ Cambridge Consultants, *Use of AI in Online Content Moderation* (Report, 2020) <https://www.readkong.com/page/use-of-ai-in-online-content-moderation-2282845> accessed 22 June 2025.

³⁶⁵ Ibid.

³⁶⁶ Ibid.

³⁶⁷ Ibid.

³⁶⁸ Ibid.

That being said, it is essential that automated moderation and content curation systems are rigorously and continuously evaluated, particularly in terms of their impact on users who are most vulnerable and at risk – such as women, especially those with public exposure. As new models are introduced, special care must be taken to ensure they do not reinforce or worsen existing societal biases that disproportionately affect marginalised groups, including women.³⁶⁹

Chapter V: General Conclusions

In the preceding chapters, I examined how the relevant frameworks of the CoE and the EU currently address the gendered dimensions of hate speech and disinformation – two distinct but interconnected phenomena that disproportionately affect women in politics, particularly through social media platforms.

Chapter I introduced the scope and impact of the issue, revealing that as women remain significantly underrepresented in politics worldwide, including in Europe, the effects of these phenomena can be seen offline, since they deter women from entering or continuing in politics.

Preliminary conceptualisation in Chapter II and legal analysis in Chapter III demonstrated that current European legal frameworks are still not adequately equipped to address gender-based hate speech and disinformation comprehensively.

Regarding hate speech, the lack of a universal definition, especially one that includes gender as a protected characteristic, continues to hinder consistent recognition across European states. As a result, significant ambiguity persists regarding the distinction between legally impermissible gendered hate speech and harmful sexist expression that, in the interest of preserving democratic pluralism and freedom of expression, may still be

³⁶⁹ Oversight Board, ‘Content Moderation in a New Era for AI and Automation’ (*Oversight Board*, 2024) <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/> accessed 27 June 2025.

protected. Broad legal definitions such as 'incitement to hatred' or 'violence' often fail to address more subtle yet equally harmful forms of gendered hate speech, such as 'negative stereotyping' or 'denigration'. This uncertainty is further compounded in the context of political debate, where greater protection is granted due to its status as a cornerstone of democratic societies.

While the ECtHR has progressively expanded its interpretation on hate speech, the treatment of gendered hate speech done in Strasbourg remains indirect, with no currently established jurisprudence. Despite this, it was made clear that – based on the Court’s hate speech standards and prior rulings on freedom of expression in political contexts –, once identified, hate speech based on gender clearly falls outside the scope of ‘legitimate political criticism’. It inherently lacks any ‘public interest’, and should therefore be excluded from freedom of expression protections. Fully recognising the phenomenon would also require acknowledging its gendered nature as a structural issue – an aspect the ECtHR has so far shown reluctance to address in cases of online gender-based violence.

Complementary CoE instruments like the Istanbul Convention offer important, albeit partial, foundations for approaching gender-based hate speech. The Convention’s lack of explicit reference to the phenomenon and narrow interpretations of potentially applicable provisions explicitly limit its utility.

Recent EU developments show more promise. Directive (EU) 2024/1385 mandates, in its Article 8, the criminalisation of cyber incitement to gender-based hatred. Additionally, the European Commission has proposed to add hate speech, including that based on gender, to the list of ‘EU crimes’. While the Directive affords Member States significant flexibility in adapting their national legislations, the adoption of the proposal would enable more specific legislation exclusively focusing on hate speech, encompassing its gendered form. Until adoption, legal fragmentation will likely persist, as only a minority of Member States explicitly recognise gendered hate speech in their national legislation.

Regarding gendered disinformation, the dedicated sections in Chapter II and Chapter III allow the conclusion that European legal and policy frameworks have yet to adequately address the issue as a distinct form of VAWIP. Although there is no universal definition, the phenomenon is understood to entail the deliberate spread of false and harmful narratives, coordinated to a certain degree, with the aim of intimidating, discrediting and

silencing women in politics. By distorting public discourse and reinforcing gendered stereotypes based on gender, the phenomenon lies at the intersection of information disorder and gender-based violence. This hybrid nature determines that existing legal tools – focused on disinformation, defamation, discrimination or hate speech – fail to capture its full scope if read in isolation. Therefore, a comprehensive, gender-sensitive legal and policy framework that recognises the phenomenon as both gender-based violence and a threat to democratic processes is urgently needed at the European level.

Ultimately, a significant share of the responsibility for addressing gendered hate speech and disinformation online lies with social media platforms, where this abuse proliferates – particularly in how they formulate their terms of service and ensure alignment with European standards.

As explored in Chapter IV, the EU’s DSA marks important progress by requiring large platforms to mitigate risks and remove content when users’ fundamental rights face foreseeable harm. While due diligence covers both hate speech and disinformation, removal obligations essentially apply to ‘illegal content’ – and its foreseeable variants, as positively established by the ECJ in *Glawischnig-Piesczek v. Facebook*.

However, by deferring the definition of ‘illegal content’ to national legislation, the DSA inevitably contributes to inconsistent levels of protection across the EU. Although hate speech is generally recognised as ‘illegal’ under national jurisdictions, the same legal fragmentation issue relating to the inconsistent recognition of its gendered form apply here. Gendered disinformation is subject to even greater regulatory gaps, as it is not even classified as illegal content and is instead addressed primarily through voluntary soft law instruments. This reliance on non-binding measures grants platforms wide discretion in how – or whether – they moderate such content.

AI technologies add complexity to content moderation. While offering tools for more effective and faster detection of gendered abuse, they also amplify abusive content via algorithmic bias incorporated in machine learning models. This highlights the need for stricter AI governance, alongside strengthened transparency reporting, that must incorporate a gender-sensitive perspective.

To conclude, while digital technologies have amplified the reach and visibility of VAWIP it is crucial to acknowledge that the phenomenon itself is not new. It is a long-standing manifestation of entrenched cultural and structural gender oppression. When gendered hate speech and disinformation surface in the political sphere, they take on a symbolic power, legitimising gendered violence as a tool to uphold the status quo. The chilling effect on female political participation not only erodes progress toward gender equality, but it also undermines the very foundations of democracy.

European frameworks must therefore recognise that tackling gendered hate speech and disinformation in political contexts is not just about protecting individuals from discriminatory harm; it is about safeguarding democracy itself. So that the myth of ‘sleeping her way to the top’ no longer echoes.

Bibliography

Literature

Agajanian T and Moran RE, 'Gender as a Central Site of Inquiry within Mis- and Disinformation Studies' (2024) 3 *Feminist Media Studies* 1.

Alcantara J and Valentim J, 'Gender-based Disinformation: A Scoping Review of the Literature, 2013-2023' (2023) 48 *ex æquo*.

Alkiviadou N, 'Hate Speech on Social Media Networks: Towards a Regulatory Framework?' (2019) 28(1) *Information & Communications Technology Law* 19.

Bardall G, 'Nasty, Fake and Online: Distinguishing Gendered Disinformation and Violence Against Women in Politics' in Gulizar Hacıyakupoglu and Yasmine Wong (eds), *Gender and Security in Digital Space: Navigating Access, Harassment, and Disinformation* (1st edn, Routledge 2022) 109.

Barker K and Jurasz O, 'Online Violence Against Women as a Key Obstacle to Gender Equality: A Critical View from Europe' (2020) 1 *European Equality Law Review* 47.

Bird & Bird, *European Union Artificial Intelligence Act: A Guide* (7 April 2025).

Bigio J and Vogelstein R, 'Women Under Attack: The Backlash Against Female Politicians' (2020) 99 *Foreign Affairs* 131.

Bontridder N and Pouillet Y, 'The Role of Artificial Intelligence in Disinformation' (2021) 3: e32 *Data & Policy* 1.

Bourdieu P, 'Symbolic Power' (1979) 4(13-144) *Critique of Anthropology* 77.

Bradshaw S and Howard PN, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation* (Computational Propaganda Research Project, Oxford Internet Institute 2019).

Brown A, 'The Who Question in the Hate Speech Debate: Part 2: Functional and Democratic Approaches' (2017) 30 *Canadian Journal of Law & Jurisprudence* 23.

Boukemia J, Sältzer M, Boyer S, 'Misogyny, politics, and social media determinants of hostile engagement against women parliamentarians on Twitter' (*Legislative Studies Quarterly*, 2024).

Casarosa F, *Freedom of Expression and Countering Hate Speech: Handbook on*

Techniques of Judicial Interaction in the Application of the EU Charter (European University Institute 2021).

Celorio R, *Women and International Human Rights in Modern Times: A Contemporary Casebook* (Edward Elgar Publishing, 2022).

Collignon S, Campbell R and Rüdig W, 'The Gendered Harassment of Parliamentary Candidates in the UK' (2022) 93(1) *Political Quarterly* 32.

Crenshaw K, 'Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color' (1991) 43 *Stanford Law Review* 1241.

De Vido S and Sosa L, *Criminalisation of Gender-Based Violence against Women in European States, Including ICT-Facilitated Violence* (European Commission, Directorate-General for Justice and Consumers 2021).

Eder N, 'Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation' (2024) 25 *German Law Journal* 1197.

Enarsson T, 'Navigating Hate Speech and Content Moderation under the DSA: Insights from ECtHR Case Law' (2024) 33 *Information & Communications Technology Law* 384.

Håkansson S, 'Do Women Pay a Higher Price for Power? Gender Bias in Political Violence in Sweden' (2021) 83(2) *The Journal of Politics* 515.

Henry N and Flynn A, 'Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support' (2019) 25 *Violence Against Women* 1932.

Holm M, 'Violence against Women in Politics: Emerging Perspectives, New Challenges' (2020) 3(2) *European Journal of Politics and Gender* 295.

Immenkamp B, Zamfir I and de Groot D, *Hate Speech and Hate Crime: Time to Act?* (European Parliamentary Research Service, 2024).

Jankowicz N and others, *Malign Creativity: How Gender, Sex and Lies Are Weaponized Against Women Online* (Wilson Center, Science and Technology Innovation Program, January 2021).

Judson E and others, *Engendering Hate: The Contours of State-Aligned Gendered Disinformation Online* (Demos 2020).

Kapelańska-Pręgowśk J and Pucelj M, 'Freedom of Expression and Hate Speech: Human Rights Standards and Their Application in Poland and Slovenia' (2023) 12 *Laws* 1.
Kasim C, 'Advancing Gender Equality: The EU's Landmark Directive 2024/1385 on Violence Against Women' (2024) SSRN 1.

Kettermann MC and Benedek W, 'Freedom of Expression Online' in Mart Susi (ed), *Human Rights, Digital Society and the Law: A Research Companion* (Routledge 2019) 58.

Krook ML, *Violence Against Women in Politics* (Oxford University Press 2020).

Krook ML and Restrepo Sanín J, 'The Cost of Doing Politics? Analyzing Violence and Harassment against Female Politicians' (2020) *Perspectives on Politics* 18(3) 740.

Krook ML and Restrepo Sanín J, 'Violence Against Women in Politics. A Defence of the Concept' (2016) *Política y Gobierno* 23(2) 1.

Lanoszka A, 'Disinformation in International Politics' (2019) 4 *European Journal of International Security* 227.

Lisinska J and Castro D, 'Why AI-Generated Content Labeling Mandates Fall Short' (Center for Data Innovation, 2024) 1.

Shifman L, *Memes in Digital Culture* (MIT Press, 2014).

Mantilla K, 'Gendertrolling: Misogyny Adapts to New Media' (2013) 39 *Feminist Studies* 563.

Martínez-Sanz R, Duránte-Stolle P and Simón-Astudillo I, 'Memes as Hate Speech: Violence, Humour and Criticism Surrounding the Image of Irene Montero' (2024) 16(1) *International Visual Culture Review*.

Matsuda MJ, 'Public Response to Racist Speech: Considering the Victim's Story' (1989) 87(8) *Michigan Law Review* 2320.

Mazur V and Chochia A, 'Definition and Regulation as an Effective Measure to Fight Fake News in the European Union' (2022) 9(1) *European Studies – the Review of European Law, Economics and Politics* 15.

Morini C, 'Countering Online Sexist Hate Speech in the European Legal Context: Between Present Commitment and Future Challenges' (2024) 104 *QIL* 17.

National Democratic Institute, *#NotTheCost: Stopping Violence Against Women in Politics: A Call to Action* (NDI 2016).

National Democratic Institute, *#NotTheCost: Stopping Violence Against Women in Politics: A Renewed Call to Action* (NDI 2021).

National Democratic Institute for International Affairs, *Tweets that Chill: Analyzing Online Violence Against Women in Politics* (2019).

Nave E and Lane L, 'Countering Online Hate Speech: How Does Human Rights Due Diligence Impact Terms of Service' (2023) 51 *Computer Law & Security Review* 1.

Ouyang X and others, 'Disinformation Reinforces Female Political Inequality and Social Misogyny' in *Proceedings of the 2021 International Conference on Public Relations and Social Sciences* 586 (Atlantis Press, Advances in Social Science, Education and Humanities Research, 2021) 247.

Parekh B, 'Is There a Case for Banning Hate Speech?' in Michael Herz and Peter Molnar (eds), *The Content and Context of Hate Speech* (Cambridge University Press 2012) 37.

Peršak N, 'Criminalising Hate Crime and Hate Speech at EU Level: Extending the List of Eurocrimes under Article 83(1) TFEU' (2022) 33 *Criminal Law Forum* 85.

Piscopo JM, 'States as Gender Equality Activists: The Evolution of Quota Laws in Latin America' (2015) 57(3) *Latin American Politics and Society* 27.

Rauchegger C and Kuczerawy A, 'Injunctions to Remove Illegal Online Content under the eCommerce Directive: Glawischnig-Piesczek' (2020) 57(5) *Common Market Law Review* 1495.

Richardson-Self L, 'Woman-Hating: On Misogyny, Sexism, and Hate Speech' (Cambridge University Press, 2017) 33 *Hypatia* 256.

Rogers R and Giorgi G, 'What is a Meme, Technically Speaking?' (2024) 27 *Information, Communication & Society* 73.

Schabas WA, *The European Convention on Human Rights: A Commentary* (OUP 2015).

Schlag G, 'European Union's Regulating of Social Media: A Discourse Analysis of the Digital Services Act' (2023) 11(3) *Politics and Governance* 168.

Sękowska-Kozłowska K, Baranowska G and Grabias AG, 'Sexist Hate Speech and the International Human Rights Law: Towards Legal Recognition of the Phenomenon by the United Nations and the Council of Europe' (2022) 35(2) *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique* 2323.

Senta C and others, *Preventing Hate Speech Online: Materials for Teachers* (Centre for Peace Studies Zagreb, 2021).

Shifman L, *Memes in Digital Culture* (MIT Press, 2014).

Sjöholm M, 'Regulation of Online Gender-Based Hate Speech and International Human Rights Law: Current Status and Challenges' (2024) 104 *Questions of International Law* 3.

Sobieraj S, *Credible Threat: Attacks Against Women Online and the Future of Democracy* (Oxford University Press 2020).

Sobieraj S, 'Bitch, Slut, Skank, Cunt: Patterned Resistance to Women's Visibility in Digital Publics' (2018) 21(11) *Information, Communication & Society* 1700.

University of Warwick and TNO, *Workshop 5: Policing of Trolling on Social Media: Ethical and Legal Issues, Deliverable 4.5* (2018).

Van Sant K, Fredheim R and Bergmanis-Korāts G, *Abuse of Power: Coordinated Online Harassment of Finnish Government Ministers* (NATO STRATCOM Centre of Excellence, February 2021).

Vasist PN, Chatterjee D and Krishnan S, 'The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-Country Configural Narrative' (2023) *Information Systems Frontiers* 1.

Wilfore K, 'Security, Misogyny, and Disinformation Undermining Women's Leadership' in Gulizar Hacıyakupoglu and Yasmine Wong (ed), *Gender and Security in the Digital Space* (Routledge 2022) 124.

Wright G, 'Dignity and Conflicts of Constitutional Values: The Case of Free Speech and Equal Protection' (2006) 43 *San Diego Law Review* 527.

Zamfir I, *Violence against Women Active in Politics in the EU: A Serious Obstacle to Political Participation* (European Parliamentary Research Service, 2024).

Internet Sources

Amnesty International, 'Diane Abbott: Violence Against Women Online' (21 March 2018) <https://www.amnesty.org/en/latest/news/2018/03/diane-abbott-online-violence-against-women/> accessed 1 May 2025.

Amnesty International, 'Nicola Sturgeon: A Case Study in Violence Against Women Online' (21 March 2018) <https://www.amnesty.org/en/latest/news/2018/03/nicola-sturgeon-online-violence-against-women/> accessed 1 May 2025.

BBC News, 'Labour MP Jo Cox murdered for political cause' *BBC News* (14 November 2016) <https://www.bbc.com/news/uk-37978582> accessed 25 April 2025 accessed 6 May 2025.

Boldrini L, 'Digital Misogyny and Character Assassination' (*Foundation for European Progressive Studies*, 24 November 2020) <https://feps-europe.eu/digital-misogyny-and-character-assassination/> accessed 1 May 2025.

Cambridge Consultants, *Use of AI in Online Content Moderation* (Report, 2020) <https://www.readkong.com/page/use-of-ai-in-online-content-moderation-2282845> accessed 22 June 2025.

Cambridge Dictionary, 'Spamming' (Cambridge University Press) <https://dictionary.cambridge.org/dictionary/english/spamming> accessed 27 April 2025.

De Naeyer B, 'Avagyan v. Russia: Strasbourg's Appetizer on Online Disinformation' (*Strasbourg Observers*, 27 June 2025)

<https://strasbourgobservers.com/2025/06/27/avagyan-v-russia-strasbourgs-appetizer-on-online-disinformation/> accessed 27 June 2025.

Di Meco L and Brechenmacher S, 'Tackling Online Abuse and Disinformation Targeting Women in Politics' (*Carnegie Endowment for International Peace*, 30 November 2020) <https://carnegieendowment.org/research/2020/11/tackling-online-abuse-and-disinformation-targeting-women-in-politics?lang=en> accessed 3 June 2025.

Di Meco L, 'Campaigning while female: How online hate holds women back' (*Politico*, 24 September 2021) <https://www.politico.eu/article/campaigning-while-female-how-online-hate-holds-women-back/> accessed 1 May 2025.

DW, 'Facebook Abuse: Top German Court Rules for Lawmaker' (*DW*, 26 January 2022) <https://www.dw.com/en/facebook-abuse-top-german-court-rules-for-lawmaker/a-60634812> accessed 21 June 2025.

Equality Now, 'The rise of deepfake image-based sexual abuse necessitates urgent and comprehensive responses from technological innovation, legal reform, and societal awareness' (*Equality Now*, 2024) https://equalitynow.org/news_and_insights/the-rise-of-deepfake-image-based-sexual-abuse-necessitates-urgent-and-comprehensive-responses-from-technological-innovation-legal-reform-and-societal-awareness/ accessed 21 April 2025.

European Commission, 'Code of Conduct on Countering Illegal Hate Speech Online' (*Digital Strategy*, 2025) <https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online> accessed 19 June 2025.

European Commission, 'Code of Conduct+ on Countering Illegal Hate Speech Online: Integration into the Digital Services Act Framework' (20 January 2025) <https://digital->

[strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online-plus](https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online-plus) accessed 19 June 2025.

European Commission, ‘Digital Services Act – Codes of Conduct’ <https://digital-strategy.ec.europa.eu/en/policies/dsa-codes-conduct> accessed 19 June 2025.

European Commission, ‘Code of Practice on Disinformation’ (*Digital Strategy*) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> accessed 21 June 2025.

European Commission, ‘Regulatory Framework for Artificial Intelligence’ (*Digital Strategy*) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 2 July 2025.

European Commission, ‘Roadmap: Hate Speech & Hate Crime – Inclusion on List of EU Crimes’ (*Better Regulation Portal*, 2020) https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12872-Hate-speech-&-hate-crime-inclusion-on-list-of-EU-crimes_en accessed 28 April 2025.

European Economic and Social Committee, ‘Freedom of Expression Is Not a Licence to Engage in Hate Speech’ (2022) <https://www.eesc.europa.eu/en/news-media/news/freedom-expression-not-licence-engage-hate-speech> accessed 12 June 2025.

European Institute for Gender Equality, *Gender Pay Gap* (EIGE 2024) https://eige.europa.eu/publications-resources/thesaurus/terms/1460?language_content_entity=en accessed 22 March 2025.

European Parliament, ‘EU Accession to the Istanbul Convention’ (*Legislative Train Schedule*, 6 June 2023) <https://www.europarl.europa.eu/legislative-train/theme-a-new->

[push-for-european-democracy/file-eu-accession-to-the-istanbul-convention](#) accessed 10 June 2025.

European Parliament, 'European Democracy Action Plan' (*Legislative Train Schedule*) <https://www.europarl.europa.eu/legislative-train/theme-a-new-push-for-european-democracy/file-european-democracy-action-plan> accessed 27 June 2025.

Fortinet, 'DDoS Attack' (*Fortinet*, 2024) <https://www.fortinet.com/resources/cyberglossary/ddos-attack> accessed 27 April 2025.

Global Disinformation Index, 'Gendered Disinformation in the European Parliamentary Elections' (10 June 2024) <https://www.disinformationindex.org/blog/2024-06-10-gendered-disinformation-in-the-european-parliamentary-elections/> accessed 6 March 2025.

Global Freedom of Expression, 'Kunast v Facebook' (*Columbia University*, 7 January 2022) <https://globalfreedomofexpression.columbia.edu/cases/kunast-v-facebook/> accessed 24 June 2025.

Global Freedom of Expression, 'The Case of Renate Kunast' (*Columbia University*, 15 October 2019) <https://globalfreedomofexpression.columbia.edu/cases/the-case-of-renate-kunast/> accessed 21 June 2025.

Goldberg E, 'Fake Nudes and Real Threats: How Online Abuse Holds Back Women in Politics' *The New York Times* (4 June 2021) <https://www.nytimes.com/2021/06/03/us/disinformation-online-attacks-female-politicians.html> accessed 6 May 2025.

Heikkilä M, 'The Viral AI Avatar App Lensa Undressed Me – Without My Consent' *MIT Technology Review* (12 December 2019) <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/> accessed 21 April 2025.

Heilweil R, 'Why Algorithms Can Be Racist and Sexist' (*Vox*, 18 February 2020) <https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency> accessed 22 June 2025.

Hope not Hate, 'The Freedom to Be Free from Abuse' (no date) <https://hopenothate.org.uk/in-search-of-hope/the-freedom-to-be-free-from-abuse/> accessed 16 June 2025.

Jankowicz N, 'How disinformation became a new threat to women' *Coda Story* (11 December 2017) <https://www.codastory.com/polarization/how-disinformation-became-a-new-threat-to-women/> accessed 17 June 2025.

Judson E, 'Gendered Disinformation: 6 Reasons Why Liberal Democracies Need to Respond to This Threat' (*Heinrich Böll Stiftung*, 9 July 2021) https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat#_ftn83 accessed 12 May 2025.

Jourová V, 'Europe Must Do More to Keep Women in Public Life' *Financial Times* (27 May 2024) <https://www.ft.com/content/418dfc3e-f113-436f-9cc3-e28df1b76ca3> accessed 30 June 2025.

Lamensch M, 'Generative AI tools are perpetuating harmful gender stereotypes' (*Center for International Governance Innovation*, 14 June 2023) <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/> accessed 22 June 2025).

Moschella M and Watts R, 'What We Learned Fact-Checking the UK Election' (*First Draft*, 19 June 2017) <https://firstdraftnews.com/joint-venture-learnings/> accessed 26 April 2025.

Oversight Board, 'Content Moderation in a New Era for AI and Automation' (*Oversight Board*, 2024) <https://www.oversightboard.com/news/content-moderation-in-a-new-era-for-ai-and-automation/> accessed 27 June 2025.

Reuters, 'Meta's Instagram Failed to Curtail Hate Speech against Women Politicians, Report Says' (14 August 2024) <https://www.reuters.com/technology/metas-instagram-failed-curtail-hate-speech-against-women-politicians-report-says-2024-08-14/> accessed 22 June 2025.

Ro C, 'Why Do We Still Distrust Women Leaders?' *BBC News* (19 January 2021) <https://www.bbc.com/worklife/article/20210108-why-do-we-still-distrust-women-leaders> accessed 3 June 2025.

UN Women, 'Digital abuse: Trolling, stalking and other forms of technology-facilitated violence against women' (*UN Women*, 10 February 2025) <https://www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women> accessed 24 March 2025.

UN Women, 'Facts and Figures: Women's Leadership and Political Participation' (*UN Women*, 2025) <https://www.unwomen.org/en/articles/facts-and-figures/facts-and-figures-womens-leadership-and-political-participation> accessed 21 March 2025.

United States Congress, *Amendment I: Freedom of Religion, Speech, Press, Assembly, and Petition* (Constitution Annotated) <https://constitution.congress.gov/browse/amendment-1/> accessed 2 May 2025.

Van de Heyning C, ‘Strasbourg’s Consolidation on Technology-Facilitated Gender-Based Violence: M.Ş.D v. Romania’ (*Strasbourg Observers*, 22 April 2025) <https://strasbourgobservers.com/2025/04/22/strasbourgs-consolidation-on-technology-facilitated-gender-based-violence-m-s-d-v-romania/> accessed 3 May 2025.

Jurisprudence

European Court of Human Rights

Arnarson v Iceland App no 58781/13 (ECtHR, 13 September 2017).

Beizaras and Levickas v Lithuania App no 41288/15 (ECtHR, 14 January 2020).

Buturugă v Romania App no 56867/15 (ECtHR, 11 February 2020).

Cengiz and Others v Turkey App no 48226/10 and 14027/11 (ECtHR, 1 December 2015).

Dalban v Romania [GC] App no 28114/95 (ECtHR, 28 September 1999).

De Haes and Gijssels v Belgium App no 19983/92 (ECtHR, 24 February 1997).

Delfi AS v Estonia App no 64569/09 (ECtHR, 16 June 2015).

Erbakan v Turkey App no 59405/00 (ECtHR, 6 October 2006).

Gündüz v Turkey App no 35071/97 (ECtHR, 4 December 2003).

Handyside v United Kingdom App no 5493/72 (ECtHR, 7 December 1976).

Incal v Turkey App no 22678/93 (ECtHR, 9 June 1998).

Jersild v Denmark App no 15890/89 (ECtHR, 9 September 1994).

Karataş v. Turkey App no 23168/94 (ECtHR, 8 July 1999).

Lawless v Ireland App no 332/57 (ECtHR, 1 July 1961).

Leroy v France App no 36109/03 (ECtHR, 2 October 2008).

Lilliendahl v Iceland App no 29297/18 (ECtHR, 12 May 2020).

Lingens v Austria App no 9815/82 (ECtHR, 8 July 1986).

Lindon, Otchakovsky-Laurens and July v France [GC] Apps nos 21279/02 and 36448/02 (ECtHR, 22 October 2007).

Morice v France [GC] App no 29369/10 (ECtHR, 23 April 2015).

Nadtoka v Russia App no 38010/05 (ECtHR, 31 May 2016).

Nikula v Finland App no 31611/96 (ECtHR, 21 June 2002).

Oberschlick v Austria (no 1) App no 11662/85 (ECtHR, 23 May 1991).

Oberschlick v Austria (no 2) App no 20834/92 (ECtHR, 1 July 1997).

Observer and Guardian v United Kingdom App no 13585/88 (ECtHR, 26 November 1991).

Oganezova v Armenia Apps 71367/12 and 72961/12 (17 August 2022).

Ottan v France App no 41841/12 (ECtHR, 19 April 2018).

Palomo Sánchez and Others v Spain [GC] Apps nos 28955/06, 28957/06, 28959/06 and 28964/06 (ECtHR, 12 September 2011).

Pedersen and Baadsgaard v Denmark [GC] App no 49017/99 (ECtHR, 17 December 2004).

Perinçek v Switzerland [GC] App no 27510/08 (15 October 2015).

Sabalic v Croatia App 50231/14 (ECtHR, 14 January 2021).

Sürek v Turkey (No 4) App no 24762/94 (ECtHR, 8 July 1999).

Tamiz v United Kingdom App no 3877/14 (ECtHR, 19 September 2017).

Vejdeland and Others v Sweden App no 1813/07 (ECtHR, 9 February 2012).

Volodina v Russia (No 1) App no 41261/17 (ECtHR, 4 October 2019).

Volodina v Russia (No. 2) App no 40419/19 (ECtHR, 14 December 2021).

Yıldırım v Turkey App no 3111/10 (ECtHR, 18 December 2012).

Court of Justice of the European Union

Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Ltd* [2019] ECLI:EU:C:2019:821.

Court of Justice of the European Union, *Opinion I/19* (Grand Chamber, 6 October 2021), ECLI:EU:C:2021:799.

Opinion of A.G. Szpunar in Case C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland Limited*, EU:C:2019:458.

Other

Renate Künast v Facebook (District Court Frankfurt, 2-03 O 188/21, 2021).

Legislation

Charter of Fundamental Rights of the European Union [2012] OJ C 326/391.

Consolidated Version of the Treaty on European Union [2012] OJ C326/15.

Consolidated Version of the Treaty on the Functioning of the European Union [2012] OJ C326/47.

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (opened for signature 4 November 1950,

entered into force 3 September 1953) ETS No 5.

Convention on the Elimination of All Forms of Discrimination against Women (adopted 18 December 1979, entered into force 3 September 1981) 1249 UNTS 13.

Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L328/55

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 (opened for signature 28 January 2003, entered into force 1 March 2006) ETS No 189.

Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention) (adopted 11 May 2011, entered into force 1 August 2014) CETS No 210.

Directive (EU) 2024/1385 of the European Parliament and of the Council of 13 June 2024 on combating violence against women and domestic violence [2024] OJ L [36],

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L277/1.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L168/1.

Official documents

Commissioner for Human Rights, Council of Europe, *Hate Speech Against Women Should Be Specifically Tackled* (Human Rights Comment, 6 March 2014).

Committee of Ministers, *Explanatory Memorandum to Recommendation CM/Rec(2022)16* (20 May 2022) including Appendix.

Committee of Ministers, *Recommendation CM/Rec(2019)1 on Preventing and Combating Sexism* (adopted 27 March 2019) CM/Rec(2019)1 including Appendix.

Committee of Ministers, *Recommendation CM/Rec(2022)16 on Combating Hate Speech* (adopted 20 May 2022) CM/Rec(2022)16.

Committee on the Elimination of Discrimination against Women, *General Recommendation No 19: Violence against women* (30 January 1992) UN Doc A/47/38.

Committee on the Elimination of Discrimination against Women, *General Recommendation No 35 on Gender-Based Violence against Women, Updating General Recommendation No 19* (14 July 2017) UN Doc CEDAW/C/GC/35.

Council of Europe, Committee of Ministers, *Gender Equality Strategy (2014-2017): Combating Sexist Hate Speech (2016)*.

Council of Europe, *Gender Equality Strategy (2024-2029)*.

Council of Europe, GREVIO, *Baseline Evaluation Report Andorra* (2020).

Council of Europe, GREVIO, *General Recommendation No 1 on the Digital Dimension of Violence against Women* (Recommendation No 1, adopted 20 October 2021).

Council of Europe, *Recommendation No R (84) 17 of the Committee of Ministers to Member States on Equality Between Women and Men in the Media* (adopted 25 September 1984).

Council of Europe, *Recommendation No R (97) 20 of the Committee of Ministers to Member States on 'Hate Speech'* (adopted 30 October 1997).

European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 'Joint Communication to the European Parliament and the Council – No Place for Hate: A Europe United Against Hatred' JOIN(2023) 51 final, 6 December 2023.

European Commission, 'A More Inclusive and Protective Europe: Extending the List of EU Crimes to Hate Speech and Hate Crime' COM(2021) 777 final, 9 December 2021.

European Commission, 'Commission Work Programme 2021: A Union of Vitality in a World of Fragility' COM(2020) 690 final, 19 October 2020.

European Commission, 'Proposal for a Directive on Combating Violence Against Women and Domestic Violence' COM(2022) 105 final.

European Commission, 'A Union of Equality: Gender Equality Strategy 2020–2025' COM(2020) 152 final, 5 March 2020.

European Institute for Gender Equality, *Combating Cyberviolence Against Women and Girls: Developing an EU Measurement Framework* (EIGE 2025).

European Institute for Gender Equality, *Combating Cyberviolence against Women and Girls* (EIGE 2022).

European Institute for Gender Equality, *Cyber Violence against Women and Girls: Key Terms and Concepts* (EIGE 2022).

European Institute for Gender Equality, *EIGE's contribution to the open public consultation 'Digital Services Act – transparency reports (detailed rules and templates)'* (EIGE 2024).

European Parliament, 'Resolution of 14 December 2021 with recommendations to the Commission on combating gender-based violence: cyberviolence' [2021] OJ C 251.

European Parliament, 'Resolution of 16 September 2021 on the EU Strategy for Gender Equality' [2021] OJ C 456.

European Parliament, 'Resolution of 16 September 2021 with recommendations to the Commission on identifying gender-based violence as a new area of crime listed in Article 83(1) TFEU' [2021] OJ C 117/12, Annex.

European Parliament, 'Resolution of 25 November 2020 on Strengthening Media Freedom: The Protection of Journalists in Europe, Hate Speech, Disinformation and the Role of Platforms' [2020] OJ C 425/51.

European Union Agency for Fundamental Rights, *Online Content Moderation – Current Challenges in Detecting Hate Speech* (FRA, 2023) 21.

Hussain A, *Report of the Special Rapporteur Submitted Pursuant to Commission on Human Rights Resolution 1997/26* (28 January 1998) UN Doc E/CN.4/1998/40.

Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians* (IPU, 2016).

Inter-Parliamentary Union, *Sexism, Harassment and Violence Against Women Parliamentarians in Europe* (IPU, 2018).

Inter-Parliamentary Union, *Women in parliament: 1995–2025* (IPU 2025).
International Media Support (IMS), *Online Gendered Disinformation and Sexist Hate Speech*, Intersectional Feminist Media Development Learning Brief No 2.

Ipsos, *Survey on the Impact of Online Disinformation and Hate Speech* (UNESCO, September 2023).

Khan I, *Report of the Special Rapporteur on Disinformation and Freedom of Opinion and Expression* (13 April 2021) UN Doc A/HRC/47/25.

Khan I, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* (7 August 2023) UN Doc A/78/288.

ODIHR and OSCE, *Addressing Violence against Women in Politics in the OSCE Region Toolkit: Tool 2 – Addressing Violence against Women in Parliaments* (ODIHR and OSCE 2022).

Parliamentary Assembly of the Council of Europe, *Recommendation 2098 (2017) on ending cyberdiscrimination and online hate* (25 January 2017) Rec 2098 (2017).

Parliamentary Assembly of the Council of Europe, *Resolution 2144 (2017) Cyberdiscrimination and Hate Speech on the Internet* (25 January 2017) Res 2144 (2017).

Parliamentary Assembly of the Council of Europe, *Resolution 2281 (2019) Social Media: Social Threats or Threats to Human Rights?* (adopted 12 April 2019) Res 2281 (2019).

Parliamentary Assembly of the Council of Europe, *Resolution 2590 (2025) on regulating content moderation on social media to safeguard freedom of expression* (adopted 30 January 2025) Res 2590 (2015).

Publications Office of the European Union, *Tackling Disinformation and Information Manipulation* (Fact Sheet, June 2024).

Šimonović D, *Report of the Special Rapporteur on online violence against women and girls from a human rights perspective* (18 June 2018) UN Doc A/HRC/38/47.

Šimonović D, *Report of the Special Rapporteur on Violence against Women in Politics* (6 August 2018) UN Doc A/73/301.

UNGA Res 48/104, *Declaration on the Elimination of Violence Against Women* (20 December 1993) UN Doc A/RES/48/104.

UNGA Res 73/148 *Intensification of efforts to prevent and eliminate all forms of violence*

against women and girls: sexual harassment (17 December 2018) UN Doc A/RES/73/148.

UN Human Rights Committee, *General Comment No 34: Article 19: Freedoms of Opinion and Expression* (12 September 2011) UN Doc CCPR/C/GC/34.

United Nations, *Beijing Declaration and Platform for Action* (UN, 1995).

United Nations, *Strategy and Plan of Action on Hate Speech* (2019).

Vienna Declaration and Programme of Action, World Conference on Human Rights, Vienna, 14-25 June 1993, UN Doc A/CONF.157/24 (Part I) (1993).