UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF POLITICAL SCIENCE, LAW
AND INTERNATIONAL STUDIES

**Master's degree in
European and Global Studies**

The impact of online hate speech on democratic quality

A case study from Europe

Supervisor: Prof. CLAUDIA PADOVANI

Candidate: BEATRICE ANTOLA

Matriculation No. 2058782

A.Y. 2023/2024

TABLE OF CONTENTS

0. INTRODUCTION

In the last couple of decades, the cyberspace increasingly came to represent a borderless extension of the public sphere, especially through the emergence of platforms acting as public fora where one-on-one, one-to-many, and many-to-many communication has been enhanced like never before. The constitution of such a unique, unprecedented space for public discourse historically holds the promise for a form of more direct and 'absolute' democracy, but this libertarian hope has soon faded away in the face of increasingly subtle and often undetectable forms of violence whose characteristics and targets seriously undermine not only online public discourse, but offline democratic quality as well. Through a case study focused on the abuse and harassment women in the public sphere experience online, particularly in Europe, I try to trace the gaps in literature and research around the phenomenon, and propose an original framework encompassing the threefold theoretical groundwork laid by digital constitutionalism, democratic innovation theory, and technofeminism. The hope is that this theoretical framework provide a useful lens through which the phenomenon can be further observed and addressed, both in academia as well as within political institutions.

Despite it being at the centre of both academic research and public debate, hate speech – both offline and online – still remains a rather vague and controversial topic, not so much in terms of its existence and problematic nature, but rather in its definition, scope, and conflict with other principles of public discourse. One such controversy undoubtedly concerns freedom of expression and its limitations, potentially infringing upon the right of disadvantaged groups and individuals to fully participate in a democratic society at all levels, especially in the political domain and more generally in the public sphere. The scholarly debate around hate speech, its regulation and its impact on both individual and groups' freedom of expression, has been ongoing since long before the emergence of digital devices and current platformisation; however, the online expansion of the public sphere and its unique characteristics in comparison to the offline world undeniably added some complex layers to the discussion.

Building on literature concerned with the issue of hate speech at different levels, my attempt is, first and foremost, to shed some light on its definition and characteristics, in the attempt to draw the line between freedom of expression and counterproductive, hateful statements. The gap

constituted by contradicting reflection is, to this day, still unbridgeable, due to the relativistic perception of the damage caused by hate speech as well as its regulation. However, some authoritative figures have taken a stance in this respect, calling on hate speech regulation as the 'lesser of two evils': while limiting freedom of expression within democratic societies should be carefully balanced against potential backfires, the pervasive and vicious nature of hate speech – especially online – potentially erodes democracy to its very roots by limiting its inclusive purpose.

The first chapter of this thesis, however, attempts to treat the issue of hate speech through an original threefold approach, with a focus on the online public sphere and the amount of hatred that women experience within it, especially when participating in social and political debate fundamental to the functioning of a democratic, inclusive society.

First, it is fundamental to consider the more than decade-long research conducted not only on existing regulation of the digital domain, but on the normative theoretical approach to it represented by the digital constitutionalist scholarship. Laying this basis is fundamental to provide a normative framework of the principles guiding the way forward in digital regulation, which should rely on constitutional principles aimed at enhancing the democratic character of the digital sphere as an extension of the public sphere, characterised by pluralism and inclusion. A literature review on the topic will be provided in the first part of this chapter. Conscious of its incompleteness, I underline its undoubtedly wider scope, which would require a whole other work to be appropriately covered. However, an accurate selection of its main developments should suffice to provide a clear framework for my subsequent reflections. Such a selection has been operated in light of the fact that, rather than the interplay of stakeholders and regulating actors within the digital environment, the addressal of the intersection of digital regulation and gender-based discrimination rests on the necessity for normative approach, antecedent to regulation itself, which frames the digital domain as one of protection and enhancement of human rights in all their forms by drawing inspiration from digital constitutionalism.

Secondly, I dedicate some space to a section dedicated to the entanglement of that same digital sphere with gender issues, which still affect and permeate the social, cultural and political landscape and are simply transposed onto the digital world. Cyberfeminism and its most recent evolutions – especially technofeminism – cannot be possibly cast aside from a theoretical point of

view, constituting a framework in which it is possible to properly analyse the political reality of gendered online interactions, especially if our main goal is to look into the necessary gender-based differences and gaps in the chosen strategies of communication given the gap in the responses, interactions and engagement.

Last but not least, I aim to propose an optimistic outlook as to how these findings may eventually foster the emergence of democratic innovations related to more participatory digital spaces, where minorities, and women in particular, are welcome to take part in the political debates within gender-inclusive political spaces. Some have already been promoted and carried out, with mixed results – but the door is always open for more. At the same time, it would be naif to ignore the deep-rooted nature of gender-based discrimination which tarnishes the public sphere, a complex and layered issue affecting women at different levels of political life, as citizens and voters primarily, but as representatives and public figures as well. Further research is needed into the various nuances and consequences of gender-based discrimination in the public sphere, laying the basis for a radical social and cultural change, as well as adequate measures and regulation.

Moving on to the second chapter, the attempt is to use this threefold approach as the lenses through which online hate speech is analysed in the peculiar case of women in the public sphere, especially prominent figures in the political landscape, both as politicians and as reporters and journalists, who shape the course of democratic innovations and their coverage and communication within the media. Starting from the theoretical debate around the conflicting rights to freedom of expression and protection from hatred, I discuss the issue from the perspective of gender-based discrimination, drawing on transnational norms regulating hate speech. In most cases, including Europe, gender is still the great absent, as most norms only address other types of discriminants, e.g., ethnicity and religion. Another major gap concerns the impact that hateful content has on democratic society at some important level, including representation and authority: most norms refer to the functioning of democracy at large, which is, by definition, inclusive of all types of people within democratic fora; however, the much more serious democratic damage inflicted by hate speech towards authoritative figures, women representatives, women journalists, and more generally women who carry a significant role in political and social terms, is hardly addressed. For this reason, a reflection on the functioning of democracy and its innovation, inclusiveness, and improvement is a necessary third element in analysing the impact that online hatred has on women that participate in the public

sphere not only as individuals and citizens, but also as authorities and representatives of their group within the political and communicative function of democracy. The peculiarity of this challenge is that its political dimension overcomes the individual and group dimension, encompassing the democratic system, its functionality and its quality at large, for pluralism and inclusion do not only affect minorities and disadvantaged groups, but the entire democratic system and its quality.

In this respect, Twitter is taken as one of the platforms where major communication flows occur between representatives and citizens, constituting an important forum whose regulation has a profound impact on the functioning of democratic and inclusive communication within the online domain. A focus on the experience of women in the public sphere who have been and are being targeted by hate speech and multi-medial hateful content extends the reflection on violence and harassment of women in authoritative workplaces to the online domain, where data shows a strong penalisation of women and especially women belonging to minorities, which victimises them as women first and, secondly, delegitimises them as authorities.

In the last section of this chapter, some space is dedicated to the measures and roles which different actors may undertake to intervene on the issue, both from the offline side of political and social forces, as well as online platform regulations and interventions which may both limit hate speech and raise awareness around it. However, most of these responsibilities are not yet codified, especially in Europe.

The European landscape is, in fact, the focus of the last chapter, which, drawing on European regulation, data, and statements provided by women in politics both in the European institutions as well as national parliaments across the European Union, tries to fill an important gap in research. While much data is available on harassment around European women politicians, and while most reports underline in particular its pervasiveness online, no comprehensive studies exist on this specific predicament. This gap exists both in literature and research, as well as – consequently – in regulation. By drawing on available literature and data, this thesis aims at contributing to fill the gap by providing some original insights.

Ultimately, the aim of this thesis is to synthesise different scholarly approaches to not only shed some light on the issue of gender-based hate speech, but also on its potential resolutions, which require a multilateral intervention. Moreover, by carrying out a critical analysis of the potential

and limitations of digital platforms, the attempt is to draw on current issues in the field to promote a positive use of the online space as a tool of enhancement of democratic and inclusive participation in the public sphere.

1. CHAPTER: A threefold approach to online violence against women in politics: digital constitutionalism, democratic innovation, and technofeminism

The intersection of three different frameworks – digital constitutionalism, democratic innovation, and technofeminism – constitutes an originale frame for the analysis of the current digital domain, its structure and norms, and its potential as a democratic innovation represented by its unique borderless and public nature. A normative approach to cyberspace that draws on digital constitutionalism and democratic scholarship lays the basis for the enhancement of its democratic potential, with a specific focus on the necessity for inclusiveness and pluralism, particularly of women, for digital democratic quality to emerge. The political characterisation of cyberspace here outlined rests on its nature as an extension of the public sphere, providing significant opportunities for gender-inclusive participation in the political and public debate, in the attempt to raise awareness and address the existent limitation to the free expression and full participation of disadvantaged groups to the democratic process online.

1.1 Digital constitutionalism

When it comes to the regulation of the digital sphere, there are multiple issues that cannot be avoided and still to this day affect the parallel research in the field. This is especially true for the field concerned with those proposals aiming at a formal recognition and, possibly, their formalisation and integration into institutional settings.

Defining *digital constitutionalism* per se has been proven time and time again extremely problematic. The conceptualisation of the term began to take shape with the advent of the modern online platform, which gave further leverage to the power and control exercised by digital transnational corporations, especially through the content moderation, the collection and processing of users' data and profiles, and the performance of public functions (Bertolini, 2022).

1.1.1 Berman, 2000: constitutive constitutionalism

The scope and focus of the concept itself has been the object of much debate. One of the first to discuss it, Berman (2000), underlined the regulatory force of the environment and architecture, rather than the measures put in place by public authorities (Berman, 2000). By creating, through software codes, *zones* which can only be accessed through specific and proven qualifications, the private entities responsible for the creation and use of platforms practically act as regulatory powers, despite their status being, especially in the US, completely different from that of institutional authorities. The dilemma is therefore clear – what is the role of constitutional values, which should lay at the basis of any form of regulation? Is there any way we can ensure that practical measures produced by anyone that is not public authority be subject to constitutional norms?

According to Berman, the issue rests on multiple limitations that are especially embedded in American political culture. The main one is undoubtedly embodied by so-called *state action doctrine*, imposing an iron separation of public and private sphere and conferring onto both very specific connotations, among which being subject to constitutional norms when producing regulatory measures, which is clearly confined to public authority – a vision that "survives both as a matter of constitutional doctrine and popular intuition" (*ivi*, p. 1278). In Berman's eyes, there are two different ways of bypassing this impasse, one more viable than the other.

The first, brought forward by different experts, but utopian from a more practical perspective, has been deemed the *incoherence critique* and it is based on the idea that "the state action doctrine is incoherent because the state always plays a major role, implicitly or explicitly, in any legal relationship" (*ivi*, p. 1279), as "all private actions take place against a background of laws" (*ibidem*). Given that private choices are firmly rooted in the possibilities and limits produced by state-created laws, "[t]here is no clear distinction between a state invasion of property interests and its inevitable role in defining those interests" (*ivi*, p. 1280: Paul, 1991). Moreover, the ability of private individual to determine who is invested with public power proves that the latter is a reflection of the dominant group's perspective, as much as the private sphere is an embodiment of "values, languages and beliefs" (*ivi*, p. 1281) derived from the same culture produced by public authorities' regulations and decisions.

According to the author, this configuration is transposed onto cyberspace as well. However, as it has been anticipated, it is way more problematic because of the extremely extensive power of private authorities in the digital realm. Unsurprisingly, "self-proclaimed cyber-libertarians have argued both that cyberspace is inherently unregulatable by territorially based sovereigns and that […] such a failure is to be celebrated because it will usher in the promise of *bottom-up* regulation created by non-state actors" (*ivi*, p. 1281: Post, 2000). However, "to the extent the "private" ordering in cyberspace depends on rules of property and contract, it is relying upon norms created and enforced by the state" (*ivi*, p. 1282: Radin & Wagner, 1999). Additionally, the abovementioned entanglement between the electorate and the state makes it problematic and perhaps impossible to trace a clear line separating public authority from the preferences and choices of private individuals.

Despite all the valid critiques concerned with the fundamental incoherence of the state action doctrine, looking at the factual developments makes them virtually pointless, as neither the populace nor the courts themselves have taken any steps backwards on the subject. There are multiple explanations provided by different experts that Berman puts forward as – possibly joint – possibilities, the most obvious one being that "the distinction, however illogical, actually captures a fundamental societal intuition" (*ivi*, p. 1288), as "an analytically compelling theory […] may be resisted as an intuitive matter" (*ivi*, p. 1289) and "successful legal arguments generally require both theoretical appeal and a fit with lived experience" (*ibidem*: Dworkin, 1986).

Therefore, Berman casts the incoherence critique aside, not as a matter of fact, but rather as a possibility to which the main actors in the digital sphere have proven to not respond positively. According to the author, a better gimmick to integrate the necessary constitutional values into the private entities' production of *de facto* regulatory measures finds its value in what he calls *constitutive constitutionalism*, whereby constitutional norms would be used to debate fundamental values, societal and political issues, regardless of distinctions of power – in short, the focus should be on its "constitutive role in our cultural life, regardless of whether that life is lived in the public or private sphere" (*ibidem*). In fact, courts themselves, by "applying constitutional norms[,] may sometimes be a superior forum for addressing divisive political issues", and "constitutionalizing a debate may encourage a more fruitful discourse in the society at large", because "law and legal procedures are at the core of American self-identity" (*ivi*, p. 1291).

Aside from Berman's analysis, many other authors have been debating the limitations of a possible "digital constitutionalism". Due to space constraints in this work, I consider it more appropriate to discuss those authors that took up with that branch of the discussion concerned with the limits of digital constitutionalism's actual institutionalisation, as that is what I am mostly concerned with, namely the possibility for public authorities to regulate the digital sphere more fundamentally, and/or for private entities to be subject to those specific constitutional values, which have marginal groups, their rights, and the democratic process at heart.

1.1.2 Yilma, 2017: the limitations of digital constitutionalism initiatives

One of the authors who more thoroughly discussed the limitations of the institutionalisation of constitutional values in regulatory arrangements is Yilma (2017), whose work has been centred around the characteristics preventing most measures attempting to regulate the digital domain from being formalised as binding, with the only exception of the Brazilian initiative of *Marco Civil* (2014). According to Yilma, five main issues affect such recommendations, especially those aspiring to a formal recognition, namely: their *fragmentation*, which prevents the initiatives from creating a unitary legal corpus formally recognised as a form of constitutionalism; their consequent *disjointed goals*; their *lack of feasibility*, in that they are mainly concerned with abstract and general values which can be hardly translated into actual dispositions, especially in such a complicated environment; their *Western subjection*, which limits their scope in what is possibly the first transnational legal environment, going as far as being classified beyond public authorities themselves as a concept at large; and their consequent *lack of engagement with the digital divide* affecting different areas of the world and their realisation of infrastructures dedicated to the development of a global cyberspace (Mann, 2018). Thus, Yilma's reflection sheds light on the necessity for a more cohesive framework, addressing the concrete and practical dimension of the digital environment, especially its pervasive entrenchment in Western values and logics, which prevents existing recommendations to appropriately address the digital divide both within the same countries and transnationally.

1.1.3 Redeker et al., 2018: five criteria to classify digital constitutionalism initiatives

A more accurate classification of these documents, relying on specific criteria which may serve as generic indications for future initiatives, has been carried out by Redeker, Gill and Gasser (2018).

In their most recent work, the authors defined digital constitutionalism as "a common term to connect a constellation of initiatives that have sought to articulate a set of political rights, governance norms, and limitations on the exercise of power on the Internet" (Redeker et al., 2018, p. 303). The collection they operated, which is obviously open to new entrances, spans two and a half decades, ranging from 1999 to 2016, and therefore counts the most relevant initiatives of our day. The criteria selected for such a collection rely on the authors' understanding of the process of constitutionalisation of the Internet, despite admittedly casting aside some important initiatives due to these methodological limitations. However, such trails are necessary to define which proposals exhibit a specifically constitutional character, as "constitutions define the fundamental rules and processes of a political community", and among "the documents explored […], these substantive values, problems and principles of constitutionalism are undeniably present" (*ivi*, p. 304). Therefore, they represent an ambition that is fundamentally distinguished from any other proposal, for "the […] core dimension of constitutional rules is a foundational and primary position within a hierarchy of legal norms" (*ibidem*).

What the authors refer to as *societal constitutionalism*, can be subdivided, building on Taubner (2012), into three main phases – coming to an agreement about a set of norms by a specific group; these same norms becoming law; reflection about this building up to achieving constitutional character (Teubner, 2012). Hence, "over time, the character of the documents [should] change from purely normative statements by a particular set of actors to [attempts of] legal codification" (Redeker et al., 2018, p. 305). While it is undeniable that most of the initiatives analysed in the article are still in their most embryonic stage, it is nonetheless significant that they "have powerful political and symbolic value to governments, the global community, and those engaged in Internet governance" (*ibidem*).

The five criteria advanced by Redeker et al. in their article define in a more formal manner the characterisation of documents related to digital constitutionalism which may aspire to an actual recognition of their constitutional value and thus to their possible formalisation. First and foremost, these initiatives should be *dedicated to fundamental political questions*, namely have an "inherently constitutional character" deemed to explore fundamental rights, articulate limits of power and "advance a range of governance norms" (*ibidem*) – in short, they should potentially function as constitutions per se. As such, they should *speak to an explicitly or implicitly established political community*, namely their scope should be clear and be directed towards a determined

group of individuals; this is usually the statement or grant whereby authors intend the digital community as global. Moreover, as constitutional matters, these documents should *aim at a formalised political recognition*, namely they should already contain the seeds necessary to be planted within a legal system, "advanc[ing] the principles they espouse into the realm of political legitimacy in some concerted way" (*ivi*, p. 306). To possibly satisfy these criteria, initiatives *need to be necessarily comprehensive*, in that constitutional arrangements "tend to articulate broad moral, philosophical and legal principles rather than a narrow policy agenda", although it is important not to "inadvertently exclude efforts that remain meaningfully comprehensive while strategically prioritising the visibility of certain rights over others" (*ibidem*). Finally, to make sense as proposals of this kind, the initiatives should *embody the views of "an organisation, coalition, state or other organised group of some kind"*, in that they should not be the narrow perspectives of singular individuals, but rather weighted and elaborated collective efforts (*ibidem*).

While Redeker et al.'s article goes deeper in analysing trends over time, to the extent that matters in this work, I am merely concerned with the democratising characteristics of digital constitutionalism and its florid production of potential democratic innovations. Although it is rightful to recognise the multiple merits of every single author I quote, I decided to focus on those aspects that are relevant to the end of this work, namely those addressing the innovative democratic potential of the digital sphere.

1.1.4 Celeste, 2018: a reconciling theorisation of digital constitutionalism

A very consistent overview of other authors has been carried out by Celeste (2019), in the attempt to reconcile different views and definitions of digital constitutionalism. His reflection starts with the consideration of the double-edged sword constituted by the impact of digital development on the constitutionalisation processes of recent years – more specifically, Celeste underlines the potential extension of individual rights granted by the emergence of a digital sphere which is transnational and allows for a larger resonance of individual's freedom of expression, while at the same time warning against the risks posed by the same development, which subjects the entirety of individual's presence online to regulation and collection of personal data at the hand of private entities which, as it should be clear by now, are hardly regulated by public authorities which should technically have such individual's best interest at heart (*ibidem*). Clearly, the issue has to do with the power conferred by the digital sphere onto private, non-state actors. Hence, digital

constitutionalism, as "the ideology that adapts the values of contemporary constitutionalism to the digital society", exists as a necessary response to the significant alterations caused by the "disruptive impact" of technological developments onto the constitutional equilibrium (*ivi*, p. 2). In short, an amplification of fundamental rights corresponds to a threat to those same rights, which needs to be prevented through a re-balancing of powers in light of the disequilibrium caused by digital innovations. Given the duty of the state to ensure that private actors respect such fundamental rights, normative counteractions have been and are still being put in place to respond to the disruption, in order to foster possibilities for the expansion of rights, limit and possibly eliminate rights violations, and strike the best possible balance among powers.

More specifically, the emergence of these counteractions raise two main questions – "why these normative responses should materialise and […] what their guiding principles are" (*ivi*, p. 5). In the first case, given that constitutionalism has always existed as an instrument, if not the main one, to ensure the protection of fundamental rights and the balancing of power, then its digital extension is just a natural corollary of its original, primary and traditional function. And again, it is contemporary constitutionalism that guides the principles of its digital branching, in the attempt to restore its own state of equilibrium through an effective governance of the emerging digital sphere (*ibidem*).

However, divergences as to how this should be carried out have emerged since the beginning of the 2000s, with the emergence of the first digital platforms which actually came to involve huge numbers of users.

Fitzgerald was one of the first authors to acknowledge the nature of information society, "which is international, intangible, non-territorial, and decentralised", and thus "requires a mixed governance structure combining private sector's self-regulation and public institutions' oversight" (*ivi*, p. 6). Hence, "the exercise of power is shared between public and private actors" (*ibidem*: Fitzgerald, 1999, 2000). Nonetheless, one main limit makes this analysis rather reductive – namely, such private entities can hardly be subjected to one or more specific state jurisdictions without causing collisions (*ibidem*).

The first author to use the term *digital constitutionalism* was however Suzor (2010), with the aim to denote "the project which seeks to articulate a set of limits on private power, with particular attention to the context of virtual communities" (*ivi*, p. 7). In accordance with Fitzgerald's work,

the author "considered the contractual framework of virtual communities as their law, and […] the limitations imposed on that framework by contract law as their constitutional principles" (*ibidem*). This is to be taken regardless of considerations about fairness, for users who participate in such virtual communities explicitly or implicitly give their consent to such principles, through the access and use of platforms which embody them. Nevertheless, in order to be usable within certain borders, such platforms necessarily fall within a state's contract law. The distinction to keep in mind between Fitzgerald and Suzor is that, in the latter's reflection, constitutional law plays a twofold role – "it can be used to determine the extent to which private actors' self-regulation is complying with the values established by the state", as well as "hav[ing] the duty to inform and lead the development of contract law". In short, constitutional law is indirectly "transferred into the regulation of virtual communities via contract law" (*ivi*, p. 8). Another major difference from Fitzgerald's approach is that Suzor not only underlined the necessary coexistence of private regulation and state law outside of a determined hierarchy, but also claimed that the latter has been losing momentum and that, as a consequence, the former has been filling the power vacuum of this centrality – attempts at democratisation and balancing of powers, according to the author, have "essentially failed". Regardless of this specification, even the acknowledgment of an existent role of state law is enough to subject this perspective to the same criticism advanced towards Fitzgerald, as the existence of a state law posing certain limitations to the range of possibilities granted to private powers necessarily implies its predominance (*ibidem*).

Finally, Celeste sheds some light on Kravas's stance, whereby "state politics is no longer able to fully regulate the complexity of a fragmented and plural society" (*ivi*, p. 7: Kravas, 2010) leading to Teubner's thesis of "a long process of emergence of civil constitutions" (*ivi*, p. 7: Teubner, 2004). In summary, while both Suzor and Kravas both acknowledge the possibility for private actors' self-regulation, the former accepts that it should still be guided by state law, whereas the latter "supports the idea that societal sub-sectors are now able to lead their own constitutionalisation process" (*ibidem*).

As one can easily tell from this review, there is absolutely no agreement on the definition of digital constitutionalism, nor is it clear which instruments should be employed to bring it to life. While *constitutionalism* and *constitutional law* per se are clear notions, their transposition onto the digital domain are to this day extremely problematic. The reconciliation proposed by Celeste is represented by a more welcoming theorisation, whereby the strict division of digital

constitutionalism as a competence of either public authorities and private actors is overcome through the acknowledgment that it is rather attached to a certain sphere of regulation – namely, the digital domain – where entities simply coexist. The concept should thus be abstracted and dis-anchored, and the same should be carried out in reference to normative counteractions emerging besides the nation-state dimension, as such norms do not emerge merely in the traditional normative context, but also within a trans-sectorial, transnational environment (*ibidem*). In fact, the process of *constitutionalisation* can, to all intents and purposes, be considered alive even when norms are not properly institutionalised or unitary, but may be stratified and overlapping – as a result, there is no necessity to draw strict lines between different initiatives and their scope or formalisation. Moreover, such initiatives may have different scopes in the traditional conception as well, as they may be not only national, but also regional and transnational (*ibidem*). Examples of this sort are the Internet bills of rights, such as the pioneering *Marco Civil da Internet* in Brazil, followed by similar advocacy initiatives in the US and Europe; as well as the ICANN's resolution mechanism of disputes, or the internal rules of commercial actors within the digital domain.

1.1.5 Padovani and Santaniello, 2018: the political paradox of digital constitutionalism and the openings for democratic innovations

Reflections on the potential *democratic innovations* implicit in Padovani's and Santaniello's article (2018) are of fundamental importance for the reflections presented in this work. In analysing the political paradox countered by digital constitutionalism, which represents a manner of preserving Internet principles when challenged, they present multiple aspects of the concept which open the door to democratising processes which may be carried out in the digital domain, involving and including more and more people in political and democratic initiatives online. In fact, the emergence of the Internet generated an increased demand for fundamental rights, to which corresponded however with a decreased practical insurance of those same rights in terms of feasible infrastructures.

In short, digital constitutionalism represents a newly emerging field where documents regulating the Internet present a political rather than technical character, based on international human rights law, for the first time transposed onto the digital domain. Their lack of traditional elements of constitutionalism – such as their roots in a constituent power, the latter's hierarchical position, a legally binding character and specific procedures to amend them – does not deprive these

initiatives for a *substantial* constitutional character, including their acknowledgement of rights and freedoms and their appeal for "an institutional arrangement enabling collective action and limiting the exercise of authoritative power" (Padovani & Santaniello, 2018, p. 297: Breslin, 2009). It is undeniable, then, that these documents exist in a constitutional ideological context, where their discursive dimension and aspirational nature adds a significant layer that cannot be ignored when it comes to democratic innovation and democratising processes in an era where the face of democracy itself is rapidly evolving and coming to encompass a significant number of inedited domains, the main one being exactly the digital sphere. The relevance to democratic innovation is twofold, as it is the use of constitutional language to analyse different actors' contributions to constitutional discourse, namely they also play a part, to some extent, to an innovative conception of processes of democratisation and constitutionalisation.

This phenomenon underlines the importance of discourse in shaping political realities, especially when it comes to involve actors external to public power in processes of constitutionalisation. In fact, shared and horizontal discourse constitutes "a precondition to achieve the establishment and institutionalisation of constitutional norms for the Internet" (*ivi*, p. 297), despite much scepticism related to the non-binding nature of many of these initiatives in terms of democratisation and protection of fundamental rights which are an integral part of constitutionalism (*ibidem*). Nevertheless, we can absolutely consider the parliamentarisation of digital constitutionalisation and its dialogic involvement of institutions of different natures as an embryonic move towards codification – in accordance with the first phase of societal constitutionalism illustrated by Teuber (*ibidem*).

And yet, some undeniably risky aspects cannot be set aside. The main one is the private entities taking up an exclusively technical role, which is perceived by themselves and by the public as neutral and apolitical. This places private entities, as well as the public and regulatory authority, in a very difficult position: if it allows them, on the one hand, to stay neutral on many difficult issues, on the other, it limits their role as actors necessary in the public debate around the regulation, democratisation and constitutionalisation of the Internet. Hence, there is a concrete risk that their commitment remains purely rhetorical (*ibidem*).

After this review of some of the main reflections on the digital domain and its political risks, opportunities and regulation, I move on to another major aspect of my reflection. The above-

mentioned research finds a major role in digital innovations allowing for a type of constitutionalisation and democratisation which is very, but not radically, different from the traditional processes we may have in mind. While the development of cyberspace as a public forum goes hand in hand with an unprecedented evolution of democracy, its unique characteristics compared to the offline public sphere make it an ambiguous but potentially innovative extension of the public space and civil society, opening the door to the online hosting of democratic innovation. While most scholarship on this latter topic still rests on the offline democratic tradition, its insights are useful to digital regulation aimed at enhancing such a democratic potential. The concept is very broad, so I will narrow it down to what interests us the most going into the second chapter, namely the role played by the gender issue and the importance of its inclusion for innovative democratic quality.

1.2 Democratic innovations

As mentioned earlier, Padovani and Santaniello underline, in their assessment of risks and purposes of digital constitutionalism, the issue of inclusive political discourse, which plays a fundamental role in shaping political reality (*ibidem*). For political discourse to be inclusive, its democratic nature needs to be brought to light and strengthened beyond the limits of current liberal democracy, encouraging the involvement of minorities through inclusive fora and innovations allowing for civic engagement to contribute to existing institutional arrangements.

1.2.1 Dalton et al., 2003: historical background

One area of scholarly work concerned with political developments that has recently come to the limelight for further discussing the issue is that concerned with democratic innovations. Democracy as we mean it nowadays is a rather young conception, emerging in the 19th century as a partial result of the Enlightenment's trust in individuals' ability to actively participate in major political decisions. Particularly, this first wave was generated from scepticism towards top-down governance, leading up to a demand for democratic reforms and more participation (Dalton et al., 2003). Both delegates and trustees lost their ground, as representatives were seen as captive to parties, and thus the so-called Progressive era provided growing trust in the competence of the electorate. New measures such as the secret ballot, the extension of the franchise to women and

the primary direct elections of senators in the US were among the major introductions of the time, accompanied by an interest in experiments of direct democracy such as legislative and constitutional referenda. In Europe, proportional parliamentary elections caused discomfort in many left-wing areas due to the fear that "the masses" were not capable of making rational choices in this field (*ibidem*).

Unsurprisingly, the tide changed significatively after the World Wars, as representatives came to be necessarily conceived as apart from parties, and institutions were once again entrusted over the masses. Corporatist and pluralist politics highlighted the role of bargaining between institutions and individuals and movements external to them, representing a compromising synthesis of the political alternations of the time (*ibidem*). The take was however questioned once again in the last third of the 20th century, thanks to university students and experts underlining once again the fundamental character of democratic participation. The Free Speech Movement in the US progressively came to include more stances during the Vietnam War, including anti-imperialist and anti-racist claims. Similar movements occurred in Europe in the 1960s and 1970s, in a more or less coordinated global claim of freedom, participation and extensive and inclusive democracy (*ibidem*). In a 1999 Symposium dedicated to the "Government of the Future: Getting from Here to There", the OECD countries focused on technological development concurrence in producing a more knowledgeable citizenry, capable of making more demands on government, possibly the most fundamental transformation of the democratic process since the beginning of the century. Nonetheless, cyclical trends have surrounded democracy since its emergence: hence, most scholars warn against the conviction that the tide cannot shift back in the future (*ibidem*). And yet, the entry of technology into the arena of political participation seems to have struck a significant balance since the beginning, further intervening in the public perception of democracy in our day as well.

Different reasons have been put forward to explain the innovative trend taken upon especially by advanced democracies in the last century. One factor may be the "underlying logic of democracy", as "participation and consensus building are [among its] essential characteristics" and, "once these values become accepted", an "inevitable pressure to expand these processes" allows "greater citizen access" and ensures the "effectiveness of democratic participation" (*ivi*, p. 8). The process of modernisation intrinsic to advanced democracy may also have played a role, allowing for the emergence of public interest groups and citizen action groups, especially after the 1960s. The encounter between a renovated social and political diversity and a general increase in citizenry's

skills and resources ended up fueling a *participatory revolution* (*ibidem*). Schumpeter's model of *elite competition* provides a valid alternative, according to which "the expansion of political actors stimulates a search for new opportunities for political access and influence" (*ivi*, p. 8), whereas a somewhat contrasting view suggests that "the congestion of the government process may stimulate institutional change", as an overload fuels the demand for adequate responses to the amount of institutional necessities and thus new access points are found (*ivi*, p. 9). Finally, some claim that the diffusion of democratic innovations among advanced democracies depends on a higher level of information sharing and more common solutions being found by international and transnational institutions (*ibidem*).

1.2.2 Curato et al., 2017: reenabling democratic innovation – 12 key findings

Among political schools of thought, democratic deliberation is one of the most debated domains, due to the abovementioned tides and various layers of scepticism advanced by experts in the field. In her work, Curato analysed some key findings which have been empirically proven to work within the realm.

Democratic innovations are undoubtedly realistic. Studies find that "deliberating citizens can and do influence policies" thanks to "citizens forums being funded and implemented variously" outside the state (Curato et al., 2017, p. 29). Moreover, deliberation is essential to democracy, since it can, among other responses, "induce agreement to restrict the ability of actors to introduce new options that destabilise the decision process" and "structure the preferences of participants such that they become *single-peaked* along one dimension" (*ibidem*). It should additionally be noted that deliberation is more than discussion due to its talk-centred focus, where "dispositional factors, such as open-mindedness, are important" and emotional interaction is to be addressed through careful institutional settings (*ivi*, p. 30). In fact, deliberative democracy involves multiple sorts of communication, with an ensemble of aspects including "greeting, rhetoric, humour, testimonies, storytelling" and so on falling under the denomination of *communicative democracy* (*ibidem*). This is especially relevant in that recognising the importance of personal narratives in political claim-making allows for the recognition of the plurality of speech cultures that can "enrich the inclusive character of the deliberative system", as different cultural contexts foster diverse "systems of meanings and norms" (*ivi*, p. 30). For these reasons, while deliberation has historically been a major component of democratic development, it still holds to this day an unfulfilled potential for

innovation, in that further efficient inclusion of minorities and disenfranchised groups within the democratic process represents, in some way, an intersectional enhancement of democratic involvement at multiple levels.

More importantly, democratic deliberation findings suggest that the practice can "temper rather than reinforce elite power" through the influence of citizens upon elite framing, additionally building capacities for traditionally marginalised groups by creating a space for ordinary political actors to participate (*ibidem*). In relation to elites, deliberative democracy offers a nuanced view of power, as the latter is ubiquitous, pervading the "very process of argumentation and communication", affecting deliberative procedures and shaping the broader policy context (*ivi*, p. 31). Within deliberative democracy, however, "participants are less partisan, using independent facilitators and/or ensuring deliberations are public" (*ibidem*). In fact, the practice includes multiple layers of power, including an authoritative power leading in a way that is receptive to the concerns of the affected public and additional actors who may strategize to advance their agendas and address inequalities (*ibidem*).

For what concerns the structure of deliberation itself, Curato underlines that deliberative democrats have rarely endorsed consensus as an aspiration, given that negotiations entail agreements on courses of action, but not necessarily on the reasons for them; deliberation should thus strive for pluralism and meta-consensus (*ibidem*). This is especially true given that participation and deliberation go together: mini publics, for one, are experiments aimed at democratising other facets of political life and deepening the quality of participation (*ibidem*).

One thing that should be noted is that deliberative transformation takes time. Wide changes early on in the process may "reflect anticipation of absorbing information and group deliberation as much as the effect of deliberation proper" and can even be partially reversed in time, as subsequent reflection produces actual deliberative preferences after the stance is achieved (*ivi*, p. 32). Counterintuitively, though, deliberation prevents group polarisation, breaking off homogeneity; plus, clarity comes to substitute polarisation, in that sharing one's ideas with like-minded people allows for the building of the confidence necessary to subsequently enter the larger public sphere. Polarisation itself is hardly found when groups are run on deliberative principles through the intermediation of a facilitator (*ibidem.*). Sticking to the topic of pluralism, Curato underlines the functionality of deliberation in deeply divided societies, "forging mutual respect and

understanding across discursive enclaves" and promoting "recognition, […] social learning, and […] solidarity" (*ibidem*).

1.2.3 Landwehr, 2015: meta-deliberation and the infinite regress

For the interests of my work, it is necessary to dig deeper into the nature of deliberation and democratic innovations which, as I have mentioned, are intimately linked with equality, equity and technological innovations' way of meddling as positive tools and negative externalities depending on the consciousness of those involving them in the process.

For this purpose, the work on *meta-deliberation* carried out by Landwehr is enlightening to say the least. His reflection starts out by bringing to light a paradox laying at the core of democracy, whereby major institutions, e.g., parliaments, do not allow for deliberative interactions and, by contrast, deliberative fora can hardly be considered fully democratic. That is to say, while in most parliaments the debate revolves around the presentation of products of argumentation rather than the process itself, deliberative fora are mostly composed through elitist procedures, as members are nominated due to expertise and their veto power (Landwehr, 2015).

The author then analyses a series of strategies to escape the paradox. Adopting a wide understanding of deliberation is one main suggestion; however, this bears the risk of hollowing out the concept. Redefining the concept in order to overcome the institutionalised principle of one person equalling one vote, on the other hand, exposes the fora to the possibility of turning elitist and technocratic. Modifying the structural nature of deliberation does not necessarily bring out the best in it, either: directing the focus away from institutions and towards the public does not necessarily entail positive effects for deliberative institutional design; whereas a compromise directing legitimacy at the system at large runs the risk of dissociating deliberation and democracy altogether (*ibidem*). Hence, the meta-deliberation process suggested by the author entails the idea to deliberate not only on substantive issues, but also on how decisions are reached at certain points within the political system itself (*ibidem*).

The issue with this possibility is that non-majoritarian decision-making is a challenge for democratic quality, as distributive decisions still require a majoritarian mandate due to expertise and competence; but telling these apart from purely regulatory decisions cannot be taken for granted, as all have consequences for the allocation of resources and opportunities. Moreover, as

different visions of distribution coexist in pluralistic societies, mere delegation can be seen as technocratic (*ibidem*). Nevertheless, some non-majoritarian bodies and processes have been found to enhance democratic quality, especially if stakeholders come from minorities: fora being placed far from competitive elections may allow for a better consideration of positions and preferences (*ibidem*). However, if we consider that the plurality of conceptions of the common good are power-driven, then the virtues of deliberation are not as likely to be morally legitimate (*ibidem*). Hence, why the central question concerns the degree to which non-majoritarian bodies should be programmed and granted decision (*ibidem*).

1.2.4 Beauvais & Baechtiger, 2016: equality and equity in democratic innovations

In this respect, Beauvais' and Beachtiger's (2016) reflections on the issues of equality and equity within deliberation are of fundamental importance. While equality is commonly defined as the sameness of humanity and individuals, equity is concerned with a distribution of power and resources based on determinant social circumstances (Beauvais & Beachtiger, 2016). To understand their interaction, we should consider the different goals of deliberation, namely legitimate decisions for and by the people bound by them, while also including epistemic and ethical criteria, such as encouraging learning different opinions and providing mutual respect. As it is anticipatable, these goals cannot be achieved simultaneously at all times, hence why they are usually distributed across different components; and the health of the deliberative system depends on the variegated and interlocked forms and actors (*ibidem*).

In Beauvais and Beachtiger's article, institutional design is considered as determinant in making deliberative fora more suitable for the twin values of equality and equity. The first aspect to take into account is, foreseeably, participant recruitment: while self-selection and random selection work in terms of equality, for they grand everyone the same chance of participating or being selected, a lack of diversity can systematically compromise epistemic and ethical goals, especially if we consider that, due to structural reasons, members of minorities and emarginated groups are less likely to propose and/or accept to participate in deliberative fora (*ibidem*).

Secondarily to participant recruitment, the nature of deliberative interactions themselves plays a major role into who and how gets their voice heard. In this respect, facilitators are fundamental elements, who may carry out their role in three main manners: "turn-taker" facilitators are rather

passive presences, compared to "designated drivers" who move the conversation along, but do not add new interpretations; and quasi-participant facilitators, who editorialise or interpret the conversation themselves (*ibidem*).

In relation to facilitation, the chosen communication format matters just as much. Deliberation is more about debate than it is about discussion: the aim is to participate in the process of producing argumentation rather than simply presenting its productions and challenging them. The problem comes up when we consider that marginalised voices are usually not as strong, due to issues of confidence and isolation; hence why it is important that communication be more supportive than adversarial (*ibidem*).

The choice of interaction also matters, to a certain extent. For instance, it has been proven that, compared to online fora, face to face interactions produce both a higher level of deliberation, as well as engagement: participants have been proven to be more involved not only in terms of the process of deliberation, but also in the aftermaths, as volunteers or political roles (*ibidem*).

One final remark made by the authors concerns the nature of deliberation in terms of outcomes: if no final agreement is required for decision-making, the deliberative quality is higher, but opinions tend to remain static. However, if a final decision is to be reached, then the issue shifts to unanimity, where consensus can easily become problematic, as it has been discussed before. Moreover, composition plays a huge role, in that homogeneous and heterogeneous fora lead to different types of outcomes in this sense (*ibidem*).

1.2.5 Wojciechowska, 2018: intersectionality in democratic innovations – towards an all-inclusive approach

These considerations all lead me to a final work, undeniably relevant to the next and last section of this chapter, namely gender issues in democracy and in respect to technological innovations. In her 2018 article *Towards Intersectional Democratic Innovations*, Marta Wojciechowska discusses the importance of the intersectional perspective in deliberation.

While intersectionality is still at an embryonic stage within democratic scholarship, its relevance in public political and social discourse is becoming more and more evident, as single categories of identity are time and time again proven insufficient to analyse and address oppression. Given that democratic institutions should be inclusive, they should be invested with "the ability to incorporate

and adapt to specific circumstances of members of disempowered groups" (Wojciechowska, 2018, p. 2). The main issue resides in the overcoming of an atomizing idea of equality, accounting for one separate category at a time. According to the author, an intersectional shift would not require a radical transformation of theories and/or tools, but it does entail a new focus within democratic innovations, which should facilitate the leadership of the disempowered and promote the diversification of the context (*ibidem*).

In deliberative and participatory democracy, inclusion is central, especially in invited spaces designed by professionals to allow citizens to make decisions on urgent public policy issues. Normally, the criterion to satisfy is that of equality, reproducing the demographic characteristics of the general public on a small scale (*ibidem*). However, a further layer should be added to understand the impasse that democratic innovations face in this respect. External inclusion, namely the ability to join the decision and thus not being left out nominally, is different from internal inclusion, which provides "an effective opportunity to influence others and thus the outcomes" (*ivi*, p. 3). Democratic innovations can facilitate the former and address the latter by including members of disempowered groups, providing them with relevant information, and strengthening their political, participatory and deliberative capabilities. Theoretically, processes of informed debate and engagement in justifications of argumentation should guarantee internal inclusion (*ibidem*).

Yet, intersectional scholars have argued in favour of a broader application of intersectionality, as a "challenge that urges us to grapple with and overcome our entrenched perceptual-cognitive habits of essentialism, categorical purity, and segregation" (Carastathis, 2016). Among the three approaches to inclusion, the first ones being a unitary approach which addresses one category at a time and a multiple one analysing several identities as static, the intersectional approach is the more thorough perspective as it cares for several categories while identifying them as dynamic and diverse. "The position and experiences of individuals on the fluid identity spectrum should in fact be of particular concern", as these members face widespread discrimination and oppression, given their higher chance of being victim of physical violence, unemployment, and so on (Wojciechowska, 2018, p. 5).

As already mentioned, recruitment is the first stage where this issue emerges: while self-selection does not account for structural, physical or social abilities, random selection excludes those

members that may not agree to participate due to external constraints, as well as unregistered, homeless or informal people and migrants. Controlled selection is thus the only way to be inclusive, but at the expense of a pure idea of equality: a relevant selection should be based on fuzzy sets that follow the democratic approach, targeting members at the intersection that would otherwise only be included in either one of the identity categories. Additionally, internal inclusion would require not only these people's participation, but also their chance to take leadership positions in the organisation, management and recruitment: seeing people similar to one's experience engaged at such levels can enhance confidence and encourage participation, thus facilitating activism and commitment. The context itself in which these events take place should be diversified and carried out especially in underprivileged areas, encouraging self-selection among the disempowered (*ibidem*).

Analogously to previous authors, Wojciechowska underlines how inclusion is made easier by facilitation which, within an intersectional perspective, should choose a mode of communication suitable for disempowered participants and helpful in their overcoming of barriers. Facilitators should thus lead a discussion, interact with participants and include them in the decision-making process by allowing them to reach self-defined goals and ensuring pluralism in association with power dynamics. In a top-down approach, this includes acting as someone whose interests and opinions reflect those of the absentees. The particular experiences of the disadvantaged, however, are likely to be alien to most facilitators due to the absence of intersectional analysis within the scholarship of democratic innovations; and training the facilitators on a unitary and/or multiple analysis easily leads to essentialist perspectives, proposing a static and unique set of characteristics which is extended and attributed to all members of a certain group (*ibidem*).

Wojciechowska further adds an additional layer to previous analysis at the level of decision-making outcomes. She notices that, in focusing on the direct decision-making results rather than the longer-term outcome, most democratic innovation acts as a consultative forum whose decision may or may not be officially carried out. Even more, "not all innovations arrive at a final decision, but those that do are exclusive towards some participants as a result [of the procedure], such as majoritarian or absolute majority voting" (*ivi*, p. 9); hence why Landwehr debates meta-deliberation.

One solution proposed by the author is the creation of enclaves, as "groups repeatedly excluded from the public sphere are less accustomed to articulating their preferences and opinions" (*ivi*, p. 9). The deliberation of enclaves in subaltern counterpublics can address the issue, creating and promoting a space where groups can discuss their ideas and strengthen their argumentation. However, these initiatives can still be prone to intragroup exclusion and domination, given that most authors do not consider how different identities interact with each other, nor do they contemplate the creation of separate deliberative groups for those experiencing international exclusion. Intersectionality and intragroup diversity is thus of the utmost importance, and the easiest way is to encourage and allow "the disempowered to take up leadership themselves and use their experience to facilitate diversity" (*ivi*, p. 10).

One final layer to the author's reflection revolves around the claim of discourse representation that inequalities in deliberative fora should be addressed by investigating whether all relevant discourses are included. These refer to a "set of categories and concepts embodying specific assumptions, judgements, conventions, dispositions and capabilities", a way of making sense of the world shared by people. Individuals engage and reflect on multiple discourses, but some may find multiple ones essential to their adequate representation, some of which are likely to be contradictory. Hence, plurality of discourse is a fundamental element of democratic innovations which aim at an all-inclusive approach which does not leave out anyone (*ibidem*).

As this section dedicated to deliberative democracy with a particular focus on its necessary inclusive nature comes to an end, I turn to the third and last section of this chapter, which is intimately interlinked with the issue of digital democracy, the necessity for inclusive democratic innovations, and particularly the clear and urgent need for the involvement of minorities in the regulation of the political aspect of technological development. Gender issues are the chosen focus of this thesis, albeit conscious that a much greater work should be carried out in respect of other topics concerned with disempowerment. I will do my best to adopt an intersectional perspective and be critical about the many points that gender issues touch for different individuals and groups. A focus on technofeminism and cyberfeminism narrows down the literature review to those themes which we need to get familiar with in respect to digital democratic innovations, digital political communication, digital constitutionalism and their relation to intersectionality.

1.3 Gender issues and intersectionality

The digital world has been discussed by many feminist authors from a gender-based perspective since its emergence. *The Cyborg Manifesto* by Donna Haraway (1985) represents one of the first milestones in the field of so-called cyberfeminism, but an ever-changing domain such as that of technology in the last decades has called into question its limitations and exceedingly optimistic stances. Feminist reflections in the digital domain have evolved hand-in-hand with feminist waves, raising awareness around the emancipating potential of cyberspace, while warning at the same time about its ambiguous nature. Technofeminism – encompassing the most recent developments in the literature – represents a more critical outlook, taking into account non-negligible factors such as intersectionality.

Judy Wajcman, one of the most influential authors within the scholarship of technofeminism, talks about feminist writing identifying "the way in which socio-technical relations are manifest in physical objects and institutions, but also in symbols, language and identities, with scientific facts and technological artefacts being treated simultaneously as semiotic and material" (Wajcman, 2010, p. 144).

1.3.1 Judy Wajcman's technofeminism overcomes cyberfeminism

Different branches of interconnected feminisms have been dealing historically with the issue of techno-scientific development, with the last wave of technofeminism criticising digital innovations more specifically (*ibidem*). The main challenge they faced was by far that whereby sex-differences in the craft and use of technology are either not inherent or irrelevant. Binary oppositions typically entrenched within the Western culture, however, remark both these differences between masculinity and femininity and have further privileged since the beginning of times the former over the latter (*ibidem*: Harding, 1986). While masculinity is traditionally associated with machines due to their cultural and historical roots, feminist technology studies (FTS) have argued for the importance of daily technologies used by women as well (*ibidem*: Cowan, 1976; Stanley, 1995). In fact, the modernist association of technology with masculinity translates into "everyday experiences of gender, historical narratives, employment practices, education, design and the distribution of power" across a global society, where technology is the driving force of progress (Bray, 2007, p. 38). It is for this reason that FTS looks at technology as a potential point of political

leverage, overcoming the gender-blindness affecting social studies in the field (Wajcman 2010; *ibidem*), in order to "develop innovative analyses of the material worlds and technology's role in shaping local and global configurations of power, forms of identity, and ways of living" (*ivi*, p. 38: Lohan, 2000).

Historically, these considerations around masculinity and technology have their roots in the late 19th century, when engineering increasingly came to define technology, downgrading the significance of technical knowledge associated with women and femininity, far from the ideals of prowess and individual achievement. Different childhood exposure to technology, opposite role models, unrelated forms of schooling and segregation of the job fuel to this day the construction of man as strong, manually able and technologically endowed, and women as physically and technically incompetent" (Wajcman, 2010: Cockburn, 1983), requiring them to give up major aspects of their femininity to be deemed valuable within the scientific and technological field (Wajcman, 2010). Technical skills and domains of expertise are thus divided between the sexes, with men being seen as having a "natural affinity" with technology (Bray, 2007). However, this is a rather Western-centric view: elsewhere, there can be found clear "mismatches between image and practice, with fractured and contradictory construction of masculinity" often coexisting and challenging typically Western associations (*ivi*, p. 42).

While labour economists justify the phenomenon in terms of differences in human capital, domestic responsibility, and employment discrimination (Wajcman, 2010: Becker, 1991), the problem resides in both different socialisation processes and equal opportunity policies. Nonetheless, the issue does not lie with women themselves: we should wonder, in turn, how technoscience can become emancipatory in spite of its deep involvement with masculinity (Bray, 2007).

In the tradition of radical feminism, for one, fundamental differences between women and men are taken for granted; however, women have been systematically controlled, dominated and abused by the men, especially through reproductive technologies. Where liberal feminists see, on the other hand, neutral technologies under a biassed power, radical feminists see the objects themselves as embedded in gender-powered relations, thus rejecting most of technoscientific development as we know it. Clearly, this runs the risk of depicting women pessimistically as mere victims of patriarchal technoscience, with no agency of their own (*ibidem*).

In a different vein, socialist feminists concern themselves with the relationship between women's work and technology, which are both "crucially affected by antagonistic class relations of production" (Wajcman, 2010, p. 147). The technological revolution is seen as a trait of a capital accumulation process, and the division of labour reflects a sexual hierarchy: the gendered nature of the class question is therefore not incidental, as social relations are materialised in tools and techniques. Yet again, women's agency is discarded, for the possibility to redesign a gender-equal technology is mostly discarded by socialist feminists (*ibidem*).

The FTS agenda is both intellectual and political, the ultimate goal being the translation of scholarship into feminist practice (*ibidem*).

By the late 1980s, second-wave feminism transformed itself in response to queer, black, postcolonial, and postmodern critiques: the epistemological break shifted away from equality to differences between women, stressing the interlinkage between gender and other axes of power (*ibidem*).

It is in this context that constructivist approaches emerged, shifting theoretical and empirical attention "from engineers' decisions to the complex social negotiations and contestations, the heterogeneity of expertise, of interest groups and of material and institutional networks": in fact, constructivism rests on the premise of the mutual shaping of institutions and civil society (Bray, 2007), reflecting the concept of *sociotechnical systems*, according to which the social and technological are an inseparable, "seamless web" (Hughes, 1986). The focus of constructivist studies was thus upstream, looking at the process of conceptualisation and use of resources involved in design, production and marketing, which excluded women systematically. However, since artefacts can be shown to incorporate "configurations of the user", including "gender scripts", the shift of attention downstream conceived customers as users engaging actively with the physical and symbolic dimensions of the artefact (Oudshoorn & Pinch, 2003). One recent constructivist approach blends the two perspectives into a focus on the "mediation junction", locating stakeholder interactions, coalitions, and contestations within the overarching contexts of regulation, which involves the state, the market, and civil society (Bray, 2007).

A more enthusiastic perspective furthermore emerged at the dawn of the digital age, with the virtuality of cyberspace and the Internet ending the embodied basis for sex differences and blurring the boundaries between human and machine and thus between male and female. In Haraway's

view (1985) then, new technologies represented a synthesis and thus an overcoming of the typically Western Hegelian dichotomy between the One and the Other, the standard and the deviation constantly attempting to reflect it, the male status and every other status that is culturally perceived as its imperfect reproduction (De Beauvoir, 1949). Elaborating on this, Haraway depicts "a new feminist imaginary different from the material reality of the existing technological order; however, the possibility and fluidity of gender discourse in the virtual world is constrained by the visceral and lived gender relations of the material world" (Wajcman, 2010, p. 148).

There are however many shortcomings to Haraway's vision, some, but not all, drawn by Wajcman herself. The latter acknowledges and praises Haraway's ability to engage with ambivalence, moving away from, but not denying, the technoscientific roots of capitalism, militarism, colonialism, racism and sexism. From her perspective, new technologies hold a potential to cause a cultural short-circuit, becoming an asset of emancipation by challenging the everlasting cultural dualism characterising Western culture and thus disrupting the dominant narrative. However, according to Wajcman, the tension between modernism and postmodernism is here undeniable: not only are there concrete, physical and existent constraint to this potential, but its theorisation relies on linguistic elements and conceptualisations which come from and fall within that same tradition. Wajcman is in fact rather sceptical of Haraway's clear-cut and a priori distinction between contemporary and earlier technologies, fundamentally denying a collapse of the distinction on which the author rests her case: that between human and inhuman and, more deeply, natural and unnatural (Wajcman, 2010: Gill, 2007). In fact, for the longest time, women's involvement was, and partly still is, only allowed as long as it is functional to the maintenance of the status quo: in the 1950s, women could finally drive a car, at the condition that it made them more efficient at fulfilling their roles of mothers and wives. To overcome the economic necessity embedded in technology, which links it to real-world systems and institutions, it is then necessary to operate an effective shift in consciousness (Wilding, 1999).

Wajcman's attempt to overcome this impasse and advance a "coherent theoretical framework that allows us to engage with the process of technical change as integral to the negotiations of gender power relations" lays the basis for the evolution of cyberfeminism into technofeminism, built and "firmly grounded in a materialist approach to social studies of technology, including its own role in [such] studies" (Gill, 2006, p. 100). Compared to previous approaches, technofeminism avoids the fetishisation and rejection of technoscience, reclaims the role of politics within STS, and

especially focuses on diversity (*ibidem*). In this respect, one enlightening work is that of DeCook, where she thoroughly analyses Haraway's work and criticises much of its utopianism, pointing out the "Western, patriarchal violence" which has catalysed technological innovation in the first place, but especially the undeniable reality that cyborgs, or anything like them, are not outside of their politics, clearly centred on whiteness and Western epistemologies disguised as universalism (DeCook, 2021). But there is a deeper, more intricate tension between cyberfeminism and the political economy of the Western paradigm, which relies on exploited labour for the mass production of infrastructure necessary to the Internet and technological development. Hence, Gajjala expresses the need for a "subaltern cyberfeminism from below", as women remain the poorest global citizens, despite aggregate-level country-specific data showing faster increasing rates of participation to the Web when given the chance – plus, at least in the US, the digital divide has been shown to depend on race rather than gender, with the rate of Internet access converging for white men and women (Gajjala, 2003).

To overcome the shortcomings of the previous wave of cyberfeminism, more than anything an exclusionary essentialism, Wajcman's technofeminism theorises a mutual shaping relationship between gender and technology, conceiving the latter within a constructivist framework as both a source and a consequence of the former; namely, its embodiment, re-perpetuating its own rigid traits in a manner affecting the entire life of a technoscientific artefact. Hence, technofeminism attempts to deconstruct the designer and user divide, emphasising the interconnectedness of all phases, as "the social meanings of technology are contingently stabilised and contestable, and thus the fate of technology depends on its social context" (Wajcman, 2010, p. 150): "both technology and gender are thus products of a moving relational process, emerging from collective and individual acts of interpretation" (*ivi*, p. 150). The hypothesis of *coproduction*, then, refers to the dialectical shaping of gender and technology, avoiding essentialism and aiming at more democratic forms of technology (Bray, 2007).

1.3.2 'Old' cyberfeminism and 'new' cyberfeminism

Some issues lying in previous forms of cyberfeminism, however, still affect the more recent waves of technofeminism. Historically, cyberfeminism has been "largely nomadic, spontaneous and anarchic" (Wilding, 1998), which has allowed for maximum freedom at the price of lacking organisations, networks and clear theoretical foundations; and while cyberspace is "a crucial point

of gender struggle desperately in need of gender diversification" (*ibidem*), the third feminist wave has failed to grasp its potential as a new space for emancipation.

To overcome the above-mentioned issues and especially the lack of a collective conception of the ideology and consequent movement, the First Cyberfeminist International was summoned on September 21st, 1997, in the attempt to foster debate around a transnational and transcultural movement. Multiple issues were discussed, including the issue of self-representation in cyberspace and the reconfiguration of networked bodies (*ibidem*). Cyberfeminist body-centred art incorporated visual and textual deconstructive projects, discussing feminine desire and sexuality. Interpolating women's bodies in cyberspace can raise multiple real-world lacunae, namely the absences and deconstructions necessary to build a "multifaceted, fluid, and conscious feminist presence" (*ivi*, p. 56).

One main problem with cyberfeminism constitutes a vicious cycle: the goal, philosophy, and definition of the approach itself is utterly unclear, and while "an epistemological and ontological anarchy that is […] open to any possibility [threads] its way […], the dogma has yet to solidify" (*ivi*, p. 56). Moreover, those "who have found their way into cyberterritories are generally those who have economic and cultural advantages in others, awarded through class, cultural and race position" (*ivi*, p. 58). The use of the Internet transforms material and "corporeal lives in a number of complex ways that both resist and reinforce [such] hierarchies" (Daniels, 2009, p. 101).

Many sociological approaches have taken up the issue: Sassen discusses "the embeddedness of the digital sphere in the physical, material world", pointing out how it "enable[s] women to engage in new forms of contestation and in proactive endeavours in multiple different realms, from political to economic" (*ivi*, p. 102: Sassen, 2002); whereas Kendall claims that it reproduces, rather than subverting, white, heterosexual, masculine hierarchies of power (*ibidem*: Kendall, 1996). Others, like O'Brien, have focused on the policing of gender identity online, which constitutes a significant obstacle to the concept of identity tourism (*ibidem*: O'Brien, 1997); and some, like Pitts, have more specifically discussed the issue of (dis)embodiment online (*ibidem*: Pitts, 2004). The general conception of its potential has changed to the point that some distinguish between the 'old' cyberfeminism, characterised by a utopian vision of post-corporeal women overcoming the patriarchal system, and a 'new' cyberfeminism, less naïve and more about confronting this new dimension of existence from the bottom-up (*ibidem*: Fernandez et al., 2003).

In practice, depending on the circumstances, engagement with cyberspace can either resist or reinforce hierarchies. In this work, I have chosen to focus on gender as the main axis, while attempting to keep an open mind, attentive to intersectional perspectives and uses of the digital domain. I now move on to the fourth and last section of this chapter, which aims at putting together this threefold theoretical basis, concerned with the regulation of cyberspace, its democratic potential and a specific focus on how these two matters are relevant for women's emancipation. Exploring some practical examples of an emancipatory use of the cyberspace, as well as data on gender biases collected online by authors studying the technical functioning of the Internet and its algorithms, I do my best to underline how an attentive, inclusive and balanced regulation of the cyberspace can protect and foster the safety of democratic spaces available to gender-related minorities attempting to find a way out from the social and essentialist constraints of the material world.

1.4 Practical uses of the cyberspace: democracy and gender emancipation

1.4.1 Pulling the strings together: digital democracy

Digital democracy is a much discussed, but rather fuzzy concept, linking practice of institution to collective self-determination, mediated through digital infrastructures (Berg & Hofmann, 2021). Its dynamics, consisting of the conflicting democratic principles of freedom, equality and popular sovereignty, create a changing landscape embracing new possibilities of imagining, realising and practising democracy, through an open-ended phenomenon connecting democratic self-government with the digital infrastructure (*ibidem*).

Historically, digital democracy has developed through three main stages, depending on both the digital infrastructure and the political momenta. In the 1980s, information flows among citizens were strengthened and participation was facilitated through the exclusion of mediators in direct-democratic "town hall meetings". In this phase, the notion of technology in democracy was that of a mere tool (*ibidem*). In the second phase, that of *virtual democracy*, the meetings were no longer within a "town hall", but rather in a "global village" which embodied the utopian idea of a denationalised democracy out of the government's reach. Besides this techno-libertarian imaginary of free and equal individuals, a more communitarian vision emerged through a lost community of civic commons, which better reconciled itself with the early 2000s' call for regulating the digital

infrastructure (*ibidem*). Finally, in the Web 2.0 democracy, or *network democracy*, netizens from the 1990s were able to turn into content producers, contributing directly to the public discourse. This lacked the utopian and revolutionary basis of virtual democracy, focusing on a "mass participation [with]in representative democracy" (*ibidem*: Froomkin, 2004) where freedom was no longer elitist, but to be reconciled in real life with "cultural diversity, political discourse and justice" (Benkler, 2006). This characterisation recalls the parallel evolution of the new cyberfeminism which, moving on from the essentialising notion of the cyborg as an all-encompassing product, took a glance at the more inclusive and intersectional perspective. The early 2000s reflections on digital infrastructure thus influenced both perspectives, albeit separately, by questioning the libertarian Western paradigm and being more attentive to diversity and the concrete political culture which was lacking such considerations. Once again, just like technofeminism, the price for mainstreaming digital infrastructure soon turned out to be "the amalgamation of commercial and emancipatory logics, with new business models driving [such] services while commodifying the private sphere and the human mind" (Berg & Hoffmann, 2021, p. 5). Socialist reflections within the new cyberfeminism recall this point, criticising the lack of intersectionality and decolonisation which condemns minorities, especially in the developing world, to the digital divide and to logics of labour exploitation within the realm of technological development.

Once it has been accepted that the digital realm could implement an evolution, but not fall completely outside of democratic institutions, some experiments have been interested in the possibility to renew and modify them. The first set focuses on their decay and destabilisation, underlining the loss of "agenda-setting power of mass media, [… replaced by] a more direct form of political communication" (Berg & Hofmann, 2021, p. 6: Dahlgren, 2005; Coleman, 2017). Such services are pessimistically seen as a threat to democracy, raising the question as to whether the latter "needs to be defended against the fragmentation and hybridisation of the public sphere, in the light of hate speech, violence and disinformation campaigns" (*ivi*, p. 6: De Blasio & Viviani, 2020; Howard, 2020; Bennett & Livingston, 2020). Such an "instrumentarian" connotation of communicative power puts the blame on technology, depicting platforms as systems driving democratic change through distortions such as echo chambers (*ibidem*: Pariser, 2011; Sunstein, 2017). A second, more optimistic set presents multiple narratives on democratic transformation whereby technology is instrumental to an active reform of representative institutions: a growing

number of civic tech organisations are rising beside legislative functions, including platform parties and social movements such as the cases of Barcelona and Madrid, where municipal platforms reunited citizens around institutions, involving them in the political decision-making process (*ibidem*). In this scenario, technology is neither seen as a mere tool nor as an independent driver of change, but rather as a "space of possibilities", which need to be located and understood within a broader constellation of social, cultural and economic factors (*ibidem*: Hubig, 2006).

Four different domains of democratic transformation can be analysed through these lenses. The first is *democratic government*, which through technology empowers citizens to engage with public administration on all levels, casting aside political parties and other hierarchical structures in favour of a more horizontal form of civic collaboration. This cannot clearly take the form of direct democracy in most instances, with organisational limits promoting a simple improvement of representative democracy through the accountability and transparency favoured by a more direct exchange between citizens and institutions, although in some cases, like the above-mentioned innovations of Barcelona and Madrid, citizens become more actively involved in consultation, to the point of actually taking part in the decision-making process (*ibidem*). This is all the more common in small-scale scenarios, setting a standard for bigger experimental practices. The hidden risk of this aspect is that of facilitating technocratic notions such as *data democracy*, whereby digital democracy turns to science and management perfecting the information base as a condition for effective policing. Clearly, this falls within the spirit of paternalistic liberalism, whereby data about the electorate is seen as a substitute for its participation and expression, turning policies into "nudges" from above (*ibidem*).

The second dimension is that of the *public sphere*, a space for opinion and will formation essential to liberal democracy, connected with communication media through public voices and user-generated content in many-to-many communication. Through digital infrastructures, this has become increasingly important in relation to elections and parliamentary decision-making, as well as the emergence of watchdog activities in response to declined trust in institutions seen as elitist. With the public becoming generative through the loop of content creation and consumption, the legal boundaries with the formal production of news are blurring, including those between the public and private sphere, information and disinformation, rational and irrational stances, and marginal and influential positions. As a consequence, democracy may be shifted towards post-

electoral and extra-parliamentary practices online, institutionalising some form of negative rather than constructive sovereignty (*ibidem*).

The third dimension is that of *political action beyond participation* and representation. Digital campaigns aim at voicing people through crowdfunded lobbying, which transcends the passive notion of audience and complements activism through the bottom-up creation of publics, including examples of civic tech activism providing the digital infrastructure for it, such as Mastodon and Bluesky. However, this runs the risk of neglecting organisational ties with institutions, turning political participation into issue-oriented, short-term action; hence why not all of these experiments qualify as emancipatory, with some stressing constitutional boundaries. *Computational management*, for one, aims to control political mobilisation through manipulative *voter surveillance*, microtargeting specific groups of voters, fuelling identity politics, fostering populist mobilisation and fostering the rise of celebrities and political demagogues (*ibidem*). A recent example of this is the American insurrection of January 6th, 2022, incited by former president Donald Trump through his Twitter profile.

The last but not least dimension to keep in mind is the *relationship between domination and rights*. Data and datafication represent new forms of power operating through a systematic collection and commodification participating in surveillance capitalism, with digital platforms embodying the organisational form of this century, acting as governors in their own right with profound consequences on democratic infrastructure, including opinion power, will formation and self-government (*ibidem*). This is not only problematic from the perspective of platforms owners and thus governors, but from that of governments as well, as the relationship between the two is complex and often paradoxical: as customers, governments mandate cooperation and oblige platforms to grant them access to the data they collect, using them as proxies to enforce law. This relationship clearly suffers from unclear consequences for oversight and accountability on both sides (*ibidem*). As already mentioned in the previous sections, however, there are also initiatives towards a democratic re-embedding of these constellations of power, with an increasing although unsystematic politicisation of the issue across national borders which defends the citizens' vulnerability by reinterpreting their fundamental rights as the normative framework to regulate platform power (Padovani & Santaniello, 2018). Additionally, mushroom initiatives towards an "Internet Bill of Rights" are seen as evidence for a grassroot digital constitutionalism, where all actors contribute to the juridification of the digital sphere (Redeker et al, 2018). Other forms of

engagement at the micro and macro level refer to the growing political engagement of IT sector employees through leaks and walkouts, and national governments addressing platform power under the claim of sovereignty, although the latter runs the risk of strengthening the nation-state as an institution rather than its democratising potential. The one way out is for digital democracy to argue, as part of a constitutional moment, for an actual pluralistic approach (Berg & Hofmann, 2021).

The connection between political talk and political participation is well-documented by studies, indicating strict similarities between online political talk and face-to-face discussions in terms of participatory influence and effectiveness (de Zúñiga et al., 2010). The ambiguous reality of blog communities, however, is that, while they "present avenues for individuals to be part of traditional political participation activities, [they] also provide new online opportunities for the exchange of political perspectives and mobilisation into action" (*ivi*, p. 37). Thus, "the Internet has the potential to provide more politically oriented expressive platforms, as well as to serve as an additional conduit for political participation" (*ivi*, p. 38). The Internet's potential to represent a unique and unprecedented democratic innovation is made clear by its ability to "promote distinct informational and interpersonal dynamics that may reinvigorate the democratic process online and offline" (*ibidem*). Most studies of online participatory behaviour fail to take into consideration that "the context provided by the Internet means that the activities take on new dimensions and forms that are at once more visual, immediate, self-selected and impersonal" (*ibidem*). If, on the one hand, the ease of using and creating new communication channels "has spawned an explosion of grassroots, bottom-up participation", the Internet may also "bring elites and the public closer together, making it easier to express views to elected officials and established journalists" (*ibidem*). By facilitating many-to-many, one-to-many, and many-to-one types of communication, the Internet overcomes multiple communicative barriers which are, by their very nature, intrinsic to traditional means of communication. However, this unique characterisation of the Internet also impacts offline political participation, with a range of online activities contributing "to the convenient coordination of in-person political activities and swift mobilisation of political activists" (*ibidem*). The costs of traditional political participation, which may be unable to reach the same audience as online fora, "may encourage a different set of people to engage in online expressive participation and open the political process to a wider range of behaviours" (*ivi*, p. 39).

In this landscape, insights provided by digital constitutionalist reflections turn out to be involved in a feedback loop with democratic innovation: the codification of constitutional norms addressing the importance of democratic inclusion within the online political sphere is still far behind in most areas of the world, and Internet bills of rights accounting for the fundamental element of the political and democratic dimension of the cyberspace as part of the public sphere are mostly non-binding. On the other hand, the sole existence of these recommendations lays the theoretical basis for a multilateral intervention to be carried out by both public and private stakeholders in cyberspace, whose attention should move beyond individual users' rights as such, acknowledging their role as citizens in the online political public sphere. The central corpus of this thesis, focused on the intersectional involvement of women in online democratic debate, draws on digital constitutionalist work on inclusion and re-framing of cyberspace to prevent discrimination and enhance political emancipation. For the aim of improving democratic quality by relying on cyberspace as a realm of extended inclusion, then, even non-binding advocacy around Internet regulation matters as a normative basis.

1.4.2 Digital democracy and intersectionality: women's participation

Digital democracy emerges from the encounter of two intertwined trends: the intensification of citizens' participation in politics as a counteraction to political distrust in institutions and the increasing digitalisation of everyday life (Asenbaum, 2020). However, like any other evolution and transposition of existing institutions, digital democracy presents significant challenges in terms of inequalities, resulting in the exclusion of marginalised groups from and within participation typical of societies marked by structural power asymmetries (*ibidem*). While feminist theory offers crucial and valuable insights against identity-based discrimination, it is usually discussed in isolation from democratic theory and digital democracy more specifically (*ibidem*). The great potential of this combination was explored briefly in the 1990s by *difference democrats* through a performative politics of presence, whereby minorities were "manually" inserted within democratic innovations through quotas and other escamotages. The unaddressed risk was however that of essentialism, namely reinforcing stereotypical thinking and trapping the individual in their body. This dilemma could have been easily overcome had attention been paid to the parallel emergence of novel identity configurations on the Internet on the part of cyberfeminism, which

proposed an array of identity performances to be integrated into democratic theory through a pluralisation of participatory sites and their respective democratic functions (*ibidem*).

While sharing the goal of inclusion with deliberative democrats, difference democracy moves away from androcentrism and Eurocentrism, acknowledging that a focus on dispassionate, fact-oriented discussion runs the risk of silencing marginalised groups, resulting in internal exclusion (*ibidem*). Unsurprisingly, it has been proven that women speak less in deliberative settings, audience members are more easily swayed by men, and while women are more likely to petition and vote, they are at the same less likely "to participate in civil society initiatives and collective action, mirrored in a dramatic gender gap in political ambition among youth" (*ivi*, p. 235). Diversity as a resource for deliberation can thus only be promoted through heterogeneous publics, enabling mutual growth and the creation of "safe spaces to develop counter discourses that challenge hegemonic narratives" (*ibidem*). In short, the composition of parliamentary bodies needs to reflect not only a pluralism of idea, but also of identities (*ibidem*). In difference democracy, "the corporeal identity of marginalised subjects functions as a visible claim for inclusion, because different bodies represent different social positionalities" (*ivi*, p. 236). To avoid the latter re-perpetuating real-life stereotypes, such as seeing women as necessarily caring and nurturing, feminist insights can be extremely helpful: Young (1990), for one, criticises a type of cultural imperialism which can be associated with such instances of difference democracy, whereby marginalised groups are imprisoned in their bodies and marked as the deviant "Other" even when their depiction is positive (*ibidem*). To stay on the same example, the idea of women as mothers is not necessarily a negative one, qualifying them as mature, sensitive figures: however, seeing this characteristics as the one unifier of all people within the category is not only likely untrue, but also essentialist and reductionist of their identity and character, which is perceived as inherently different than men's qualities.

Cyberfeminist discourses, on the contrary, promote diversity through the digital performance of marginalised identities in digital counter-publics, exploring the liberating effects of anonymity as well. Donna Haraway (1985), one of if not the first cyberfeminist author, points like Young to the construction of binary categories inherent in the Western paradigm which implicitly defines one element as inferior to the other; however, while Young's difference aims at equalising the value of both elements in parallel, Haraway proposes "the metaphor of the monstrous hybrid breaking out of the dichotomous thinking and leaving the riddle unresolved: the irony of the cyborg is

constituted by its plurality dissolving unity into contradiction, a synthesis which overcomes and transcends binarism itself" (*ivi*, p. 238), albeit unclearly. In exploring this obscure depiction Haraway does of the cyborg, Turkle writes that the hybridity between human and machine depends on "a link between who we are and what we have made, between who we are and what we might create, between who we are and what, through our intimacy with our own creations, we might become" (*ivi*, p. 8: Turkle, 1984). In fact, computers simultaneously appear as objects and as thinking subjects, which humans can employ as tools to create their own individual worlds, which, however, affect them in a move of reciprocal affectivity. In a few words, computers function as mirrors for the individual self as a reconfigured other, not serving a vain mote of self-love, but rather the anxious research for the self.

In relocating the focus on the forgotten contribution of women in computing, starting with Ada Lovelace, Plant (1997) underlines that to this day women mostly execute repetitive work to fulfil plans fully developed by men; but she especially underlines the intimate relationship between women and machines, both abused and objectified by men. However, her position has been rightfully criticised for its techno-determinism and essentialist reproduction of stereotypical gender roles, setting women's agency aside.

As previously mentioned, more recent cyberfeminist work attempts to rehabilitate the strain itself by being more critical of the patriarchal and capitalist characterisation of digital spaces. In the attempt to rehabilitate cyberfeminism, Brophy (2010) resorts to the concept of *liminality*, drawing on Grosz's *in-betweenness*: the new subject constitution occurs "through the digital as a change in the perception of reality both online and offline" (*ivi*, p. 241). Additionally, Daniels (2009) draws attention to re-embodiment on the Internet and the construction of alternative body images challenging the hegemony within subaltern counterpublics. The notion of the cyborg is thus taken further and beyond online identities, "grasping the offline reconfigurations of embodiment through a digitally altered perspective of the self" (*ibidem*).

The common ground between difference democrats and cyberfeminists is that they contest gender-powered asymmetries by focusing on particular modes of identity performances, each in their own field. Cyberfeminist perspectives allow for the conceptualisation of a politics of presence online, supplanting the difference democratic notion of physical presence; however, it also draws attention to the value of anonymity, not meant as hiding identity, but as a simultaneous creation and negation

of identity constituting a specific type of identity performance, contributing to the overarching goal of inclusion in some instances.

While difference democrats developed their accounts of inclusion by relying on three main types of political institutions – namely, representative state institutions, citizens' participatory institutions, and social movement organisations – proposing continuous performances for all of them, an intersection with cyberfeminism would suggest that different institutions fulfil different democratic functions, which are "best facilitated by different modes of identity performance" (*ivi*, p. 242). Drawing on the concept of participatory spaces within feminist studies, the author distinguishes the same three types of institutions as *empowered* spaces, *invited spaces* and *claimed spaces*, the crucial differentiation being informed by the social positionality of the creators of the respective spaces and the actors within them. The metaphor of space translates the notion of physical relationality into social relationality: space as a relational concept makes imagining a discursive power architecture possible, with participatory interfaces being carefully constructed to challenge the asymmetrical power relations inscribed in societies.

Empowered spaces "fulfil the democratic function of accountability to create a strong link between citizens and their representatives" (*ivi*, p. 244); this is facilitated by *identity continuation*, or the "performative enactment of physically embodies identity as advocated by difference democrats" (*ibidem*). Online, this is carried out through "the digital replication of analogue identity performances" (*ibidem*). Empowered spaces, such as democratic fora legitimised by the citizens to make political decisions on their behalf, are expanding online, and these continuous identity performances "serve the democratic function of accountability, connecting representatives to their constituencies", as "one of the prime goals of digital government policies is to increase transparency" while also "strengthening the bond of constituents with their representatives" (*ivi*, p. 246). Besides, legislators' digitised identity performances facilitate publicity and strengthen interactivity, in a two-way street that also comprises civic feedback, allowing the inclusion of opinion in governmental decision-making. Their websites are completed by their social media presence, creating a networked self in accordance with feminist media scholars: for instance, the 19.4% women MPs in the UK make up the 29.4% of the total MPs who actively engage on Twitter, which gives credibility to the cyberfeminist theory of women's empowerment through online engagement (*ivi*, p. 247).

On the other hand, *invited spaces* "are designed to facilitate inclusive dialogue and are oriented toward the ideal of freedom from domination" (*ivi*, p. 244), allowing citizens to encounter each other as well as state agents. Digitally enabled *identity negation* corresponds to this logic, facilitating identity expression and making a discursive rather than visual diversity claim. Invited spaces online thus hold the potential to allow for an inclusive processes of citizens' participation, where identifiability proves significant in that it can be temporarily negated, countering prejudice and discrimination, as well as levelling the playing field and promoting meritocracy. Some cases illustrate the nature of difference democrats' counter-publics as enclaves of deliberation for disadvantaged groups, while at the same time exemplifying "a cyberfeminist practice of peer support", as anonymity can "work against the chilling effect and alleviate peer pressure and concerns about surveillance" (*ivi*, p. 248). Unsurprisingly, "the effect of diminished participation through enforced identification comes along with significant gender differences" (*ibidem*): compared to men, the engagement of female participants dropped significantly when public identification is required.

Finally, *claimed spaces* flourish on *identity exploration* "to facilitate expressive freedom" (*ivi*, p. 246), corresponding with social movements' need to communicate with the public: to be heard in the media, they need to "form colourful and innovative collective identities and forms of protest" (*ibidem*).

The feminist issue within digital democracy is explored by Marie Tessier in her book *Digital Suffragists* (2021), where she talks about the need for further measures to ensure women's rights online, as they are not seen as credible interlocutors, let alone legitimate leaders in the digital world networks constituting public life. According to her account, the design of apps and the terms of access on social media platforms "perpetuate women as the silent sex due to two factors: bias in design and sexism online" (Franklin, 2021, p. 1799). This is presented as "symptomatic, if not causal, of the physical and psycho-emotional violence that women face in everyday life and public office" (*ivi*, p. 1799). Despite the many limitations of her work, most of all Western-centrism, her conclusions according to which the world needs 'systemic change and action on many levels' to tear down the obstacles in the way of full suffrage are undeniable (Franklin, 2021).

Over the past 30 years, many major organisations, especially the US-based non-governmental organisation National Democratic Institute, have put an effort into increasing women's

participation as voters, political leaders, elected officials and activists around the world. However, evidence has shown that "there is a defined digital gender gap, which prevents all women from having equal access to technology, particularly in developing countries" (NDI, Ipsos, 2014, p. 3). In fact, one major research gap concerns "the distinct opportunities and barriers for women using technology as voters, elected officials, candidates and activists in developing democracies" (*ivi*, p. 3). What existing literature does cover, however, is the "general implications of digital technology usage for political participation, activism and empowerment", uncovering the "general challenges to achieving equal access for women to online and mobile technologies" (*ivi*, p. 5). According to the available data, where women do have access and ability to use technology, this "expands opportunities to seek out information, facilitate communications among diverse groups of people and enable the organisation of coalitions around particular issues"; this engaged use of technology is "particularly [relevant] for marginalised groups" (*ibidem*). However, the gender gap is significant especially in the most vulnerable and disenfranchised areas: in low and middle-income countries, women are 21% less likely to own a mobile phone than men; and the more general Internet disparity signals that 16% fewer women use the Internet (*ibidem*). Nonetheless, the Internet gender gap is pronounced in both developing countries as well as developed countries, including Italy, with a 10% difference.

Unsurprisingly, the main obstacle is the cost of access to digital technology in areas of the world at an economic disadvantage, but this is especially acute for women who, due to social and cultural factor, often lack control and access to household finances, or whose salary is not sufficient to purchase technological devices for themselves (*ibidem*). Another challenge is traceable in technological education, because science and technology are often considered male domains, and thus disadvantaged women hardly receive adequate training (*ibidem*). The lack of popular websites in local languages also represents a challenge for uneducated women, especially given that a disproportionate number of the world's illiterate population is female (*ibidem*).

Programs that use digital technology may require gender mainstreaming initiatives, defined as "a strategy of promoting gender equality by ensuring that gender perspectives and attention to the goal of gender equality are central to all activities, such as policy development, research, advocacy/dialogue, legislation, resource allocation and planning, implementation and monitoring of programs and projects" (*ivi*, p. 6). This would require "a consideration of the needs, experiences, interests and technology access levels of women throughout all phases of the program" (*ibidem*).

Several international organisations have developed strategies and guidelines for digital technology policy: the UN Women, for one, "aims to build better infrastructure and developing women's technology skills", while the World Bank tends to be more policy-based, and the International Telecommunication Union collects gender-disaggregated information and documentation on women's usage of their tools (*ibidem*). Within the political domain, the major uses that global citizens make of technological tools include communicating with others, sharing information and accessing information (*ibidem*). Specifically political engagement includes participation in online discussion or networks and promoting political campaigns (*ibidem*). In the workplace, efficacy can be enhanced by using technology to coordinate and organise meetings (*ibidem*).

The high costs of Internet and mobile are directly cited by most study participants in their work, hindering their access to technology, while a smaller but significant group faces sociocultural barriers, including "discomfort with using digital technology in public and the threat of censorship or digital surveillance" (*ivi*, p. 14). When discussing the forward, participants cite "lower access costs, training courses on how to use specific types of digital technology and resource centres at which they can use these tools" as key methods of overcoming the barriers, highlighting the need for "both incremental programmatic initiatives and long-term structural change" (*ibidem*).

For a democracy theory of change that is intersectionally inclusive of women in light of cyberspace as a branch of civic society and public sphere, the individual, institutional and socio-cultural aspects of political participation need to be accounted for. First and foremost, it needs to be ensured that available programs do not "reinforce existing gender norms or further marginalise women from political engagement", carefully selecting the technological tools used for the program and the way it is used for its activities (*ivi*, p. 16). Moreover, blended approaches are necessary to build technology skills into women's political participation activities, including the use of technology centres where possible to allow women to adequately access these tools, as well as integrating a gender perspective into ICT-related activities (*ibidem*). Where technology is proven to further the project's goals, it is fundamental to prioritise and build in resources for an ongoing monitoring and evaluation of the course, as to carry out the adequate reassessments where needed (*ivi*, p. 17).

1.5 Analytical framework and methodology

Looking forward, the attempt is to rely on this threefold approach as a lens through which more complex and specific issues can be observed, analysed and discussed.

The focus on digital constitutionalism, for one, serves as the basis for a normative approach to the regulation of the digital domain, not only in terms of what goes on within it, but also in terms of the potential limitations of technology as a wider topic, how such limits can be addressed, and whether their regulation through a multilateral intervention of actors and stakeholders can turn the issues into democratic and inclusive improvements. It is through these lenses that we can shed light on existing measures regulating or recommending a safer and more positive use of the digital domain, especially digital platforms. While not all existing norms present a focus on hate speech, especially in relation to gender-based discrimination, their principles and aims can be integrated into a wider normative approach, potentially addressing those aspects of cyberspace which lend themselves to hosting and promoting inclusive and democratic values and uses, especially of digital platforms.

These values, historically embodied in cyberfeminism and, subsequently, in technofeminism, constitute a significant gap in regulation and thus in the organic structure and potential use of digital tools. While the optimistic premises of the cyberfeminist literature, depicting the technological and digital domain as a potential for women's social and political emancipation, have partly been disappointed by both the design and the use of digital tools, there is nonetheless a wide margin for both public and private stakeholders to bring to life the potential enhancement of inclusion and democracy that the existence of a borderless social and political forum like the Internet represents. This controversial nature of the online domain is especially visible when it comes to women who hold socially and politically authoritative positions, representing their group in the fundamental and impactful context of the public sphere. For this reason, I focus my research on the potential and dangers these women encounter when carrying out their public duties online, communicating with citizens and exposing themselves as public figures.

The political and social role these women have online and offline cannot be extrapolated from the context of a democratic society within which they act as citizens, representatives, and individuals belonging to a minority. Unsurprisingly, intersectional identities constitute a higher risk for women in politics, who are more frequently targeted by online hatred, as shown by multiple studies. Hence, leaving out the impact the phenomenon has on democratic quality at large would imply missing

out on the more significant consequences it has not only on individuals, nor groups, nor institutions, but on democratic societies at large. While most research on democratic quality and democratic innovation only marginally refers to the significance the additional area of the public sphere constituted by the online domain, and does not directly discuss the dangers that hate speech poses to democratic inclusion, I consider the contradicting nature of the cyberspace to embody an important spiral into further research on its positive and negative potential for democratic quality. Its innovative nature, different from any other antecedent public space, undoubtedly requires further research within the field of democratic innovation, with a focus on democratic inclusion, minorities' freedom of speech, and their right to safely participate in the democratic process.

Pulling all these strings together is not an easy task, especially when there is a complete gap in literature addressing the issue of online hate speech towards women in politics from these perspectives. While much data exists on both online hatred towards women and the discrimination of female public figures within political institutions, the two diagrams hardly intersect in existing literature, despite many studies underlining the pervasiveness of online violence towards women in politics and its vicious, yet hidden consequences. While this specific type of phenomenon is recognised as intersectional, for it cannot be merely considered violence against women, nor merely violence against public figures, it has not been explored nearly as much as it should be to address an overwhelming phenomenon which calls for public and private regulation, deeper awareness of the extension and consequences of gender-based discrimination in the public sphere, and the impact it has on democratic quality.

The present work cannot possibly cover such a gap, but the hope is, through this threefold approach, to gather the available data and give it a new meaning, encompassing all these aspects. Pulling the strings of these three antecedently isolated scholarships constitutes a framework I consistently refer to throughout the entire work, in the attempt to forge an original perspective on the digital domain as we know it, its dangers and limitations, and its potential for a pluralistic, gender-inclusive democratic enhancement. Starting from the social and cultural root causes, which affect women's participation in the public sphere and political domain, I rely on technofeminism to address the transposition of such issues into cyberspace. Linking this gender-sensitive perspective to democratic quality comes rather naturally, as gender-inclusive participation in the public domain does not merely affect women as individuals or as a group, but determines the plurality of voices and inclusion of perspectives that lay at the heart of a healthy democratic

society. From a normative perspective, digital constitutionalist literature comes to the rescue through its vision of a re-framing of cyberspace based on constitutional principles, which traditionally refer to elements determinant to democratic quality such as inclusion, equal opportunity, transparency, accountability, and so on. More recent developments in the field have specifically touched upon gender-related issues, linking the digital constitutionalist vision with gender-inclusiveness (Bertolini, 2022).

For this work, I have mostly relied on existent literature and data reports concerning the impact of online and offline abuse and violence against women in the public sphere. Multiple studies have been carried out by both international institutions such as the United Nations and the European Union, as well as grassroot and non-governmental organisations, think tanks, and individual researchers: however, most of these are concerned with either online violence against women in general, or offline violence against women representatives and journalists, meaning the intersection between online violence and institutional violence is mostly left untouched. Relying on different types of sources sheds light on the issue at different levels: while official data reports are mostly limited to keeping track of the percentages, grassroot organisations discuss the root causes and recommend some possible paths forward, whereas academic research provides a theoretical framework. In some cases, women representatives have been directly involved through interviews and have had the chance to express their feelings and ideas in the aftermath of targeted attacks. Accounting for all these different perspectives allowed me to paint a 360-degree picture of the issue, in the attempt to resort to the root causes of the phenomenon and show how no crucial change is possible if the traditional paradigm shaping public debate is not questioned according in light of the threefold frame I have adopted. Running the data and phenomena through this threefold system gives a new meaning to existing research, tracing back the issue to a normative matter, that should account for democratic and gender-inclusive principles. As a consequence, the recommendations I provide draw partly on those set out by previous research, but also attempt to resort to the original problem of re-framing the general structure of public debate, with a focus on its online realisation and the unique characteristics of cyberspace, which is both borderless and, at the same time, determined in its structure by the interest of intermediaries, users, etc.

The democratic enhancement of such a complicated reality cannot be attributed exclusively to singular entities, for the phenomenon involves intermediaries, as well as political forces, and civil society and its values at large. For this reason, building an all-encompassing framework is

necessary for a qualitative analysis of the context at all levels, providing insights and recommendations that range from the crucial change of the political and social paradigm in culture, to technical and political measures protecting women in institutional settings.

2. CHAPTER: Freedom of expression, hate speech, online violence and the consequences for democracy: reflections from the gender perspective

Digital constitutionalist reflections, meddled by techno-feminist and reflections on the potential nature of the cyberspace as a democratically innovative domain, lay the groundwork for more practical reflections on the use of the digital domain as an enlargement of the public sphere, borderless and unique in its characteristics, which is filled with promises, as well as dangers. Starting from the theoretical debate around the conflicting rights to freedom of expression and protection from hatred, I discuss the issue from the perspective of gender-based discrimination, drawing on transnational norms regulating hate speech. Reflections on the limits of absolute freedom of expression and its consequences on minorities precede the shifting of the debate onto the digital domain, but the latter's emergence undoubtedly poses a major question, given its outreach, audience, and borderless involvement of citizens and individuals of the world. Drawing on these insights, we shed some light on various stances, through which we can observe existing data on abusive language against women online, particularly women that embody a public authority, such as journalists and politicians. While women in the public sphere are only one out of many social groups which experience online abuse in the name of free speech, reflections drawn from this case study represent an opportunity to address the possibility for gender-based digital regulation, which may be extended to the protection and involvement of other disenfranchised groups, in order to fulfil the digital potential for an inclusive and democratic extension of the public sphere.

2.1 The debate around freedom of expression

The debate about the right to freedom of expression and its limitations is much older than its digital declination, although the infinite possibilities offered by the World Wide Web significantly heighten the tensions when it comes to the defence and protection of minorities in everyday speech.

Freedom of speech per se is explicitly protected in most Western countries, but so is the condemnation of hate crimes, which encompass verbal hate speech. In the US, for one, the premise

behind freedom of speech is to "protect the ability of the people, as a collective, to decide their own fate and permit true self-determination" (Downs & Cowan, 2012: Raikka, 2003). However, together with individual autonomy, the Bill of Rights also protects collective goals such as equality (Massaro, 1991). The debate thus encompasses constitutional and philosophical factors, which further result in practical issues, as both sides carry potential social harm and face significant opposition (*ibidem*). One the one hand, it is important to notice that absolute freedom of speech opens the door to hateful speech, which subordinates and oppresses members of vulnerable groups, as language has been used for the longest time to degrade, terrorise, and humiliate minorities (Downs & Cowan, 2012: MacKinnon, 1993); however, many support the idea that, on the long run, absolute freedom of speech benefits society at large, according to the idea that such a right has also significantly advanced the causes of minorities and their movements to acquire rights and recognition (Downs & Cowan, 2012: Cole, 1996).

Despite the apparent opposition, however, the two perspectives are not as dichotomous as they seem, as both speech and equality are integral aspects of the interpretive community: while communication requires an (almost) unanimous agreement around the conventional meanings and interpretations of words and figures of speech, it has also often times been used to isolate minorities (Delgado & Stefancic, 1994). This acknowledgement poses the most significant problem of the debate: for an impartial judge to make a fair decision concerning the regulation of speech, they would need to stand outside of their own interpretive community, which is clearly paradoxical (*ibidem*). That is mainly because speech is paradigm-dependent, and to think that we can somehow escape or circumvent our own internalised narratives constitutes an *empathic fallacy*: while we may be willing to overcome the limitations of our culture and experience, our own dialogue with a previously disenfranchised and isolated community rests on the dominant current regime, and thus meanings will be heavily weighted in favour of the status quo (*ibidem*). If we add to the mix the fact that the judge's knowledge is based on existing law, we need to acknowledge that the latter involves and interpretive community of its own, which has established it as the canon; but the point of the establishment of canonical ideas is precisely to resist attack, as it causes the main community to fail to hear external proposals, translating the message inaccurately or deprecating it as extreme or politically motivated (*ibidem*). In fact, if one places at the centre of their belief system the idea that all language should be absolutely free, then any type of proposed regulation will be deemed unjust (*ibidem*).

If that is so, then why have most European countries adopted quite stringent legislation to counter hate speech (Erjavec & Kovačič, 2012)? While freedom of speech per se is considered to be absolute, action is not granted the same status: thus, verbal affirmations that lose their intangibility due to their context possibly leading to harmful acts can be rightfully regulated in most jurisdictions (*ibidem*). The protection of freedom of expression, in fact, does not necessarily entail political tolerance of hate speech, as it has been proven by Australia, which "does not possess an explicit statutory or constitutional free speech protection" and has thus provided fertile ground for promoting and testing hate speech laws (Gelber & Stone, 2007). Nonetheless, the information available clearly shows that much more needs to be done, especially in critical areas like the digital sphere.

Starting from the root communitarian issue, one interesting notion that cannot be cast aside when debating hate speech regulation is what feminist Nancy Fraser defined *subaltern counterpublics* in her groundbreaking work. Through this concept, Fraser deconstructs the Habermasian ideal of the *public sphere*, which she shows to be limited to the bourgeois conception of democracy. Habermas's idea of the public sphere emerged in early modern Europe as a "body of private persons assembled to discuss matters of public concern or common interest" (Fraser, 1990, p. 74), counterweighting absolutist states by holding them accountable to society via publicity, namely requiring their functions to be accessible and guaranteeing free speech, free assembly and, eventually, parliamentary institutions (*ibidem*). On one level, this rationalised political domination; on the other, it designed a specific kind of discursive interaction, an unrestricted discussion of public matters where private interests were to be cast aside, status inequalities bracketed and discussants were to deliberate as peers in order to generate a "public opinion", namely a strong sense of consensus about the common good (*ibidem*). By rooting his notion of the public sphere in civil society, which had and partly still has its base in bourgeois men, Habermas failed to examine other non-liberal, non-bourgeois notions of the public sphere: the host of independent publics that emerged in time and that distinguished themselves through "alternative styles of political behaviour and different norms of public speech" have thus been characterised by a certain degree of conflict with the public sphere (*ivi*, p. 61). These counter-publics function both as "spaces of withdrawal and regroupment and as training grounds for agitational activities" directed towards the public at large, meaning that the final objective is "for members of different publics to talk across lines of cultural diversity" (*ivi*, p. 68-69). The vision of democracy based on

the cross-public dissemination of subaltern counterpublics and the public sphere is the only possibility to overcome the above-mentioned limitations of the communitarian paradigm around freedom of speech, which would otherwise cast aside the revindications of minorities concerned with hate speech: what Fraser calls a "post-bourgeois conception of democracy" foreseeing the coexistence and exchange of a range of publics (*ibidem*).

Unsurprisingly, then, differences have already been found by previous research in the perception of hate speech across the spectrum of gender. In general, men have been shown to rate the importance of freedom of speech significantly higher than women, whereas the latter scored higher than men in the perceived harm of hate speech (Downs & Cowan, 2021). In Down and Cowan's study, particularly, gender was shown to be the major predictor, accounting – together with liberalism – for the most variance in the perceived harm of hate speech, as counterposed to individualism, which acted as a significant predictor of the perceived importance of freedom of speech in both men and women (*ibidem*). Separate knowing – namely abstract analysis, objective observation, and the comprehension of great ideas – and authoritarianism were related to freedom of speech among men, but not among women, whereas higher intellect was shown to be related to the perceived harm of hate speech exclusively among women (*ibidem*). The study shows that what is likely to be a different degree of empathy towards victims of hate speech determines a different perception of its harm based not only on general attributes, but also on gendered ones (*ibidem*).

2.2 Hate speech: definition and existing regulation

In recent years, several sources have indicated that bias-motivated violence and hostilities are on the rise (Massaro, 1991). Aside from defendants of free speech and condemners of hate speech, a nuanced grey area has emerged in legal scholarship, attempting to promote tightly worded, cautiously progressive measures proscribing targeted vilification of minorities (*ibidem*). In 1952, for instance, the US Supreme Court sided with hate speech regulation due to the perceived interest of the State "in preventing violence and protecting the esteem and dignity of its citizens" (*ivi*, p. 212). The Council of Europe's Protocol to the Convention of Cybercrime, for its part, defines hate speech as "all forms of expression that seek to spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against

minorities, migrants and people of immigrant origin" (Carlson, 2017: Van Blarcum, 2005, p. 781). The Draft Online Safety Bill, published recently by the UK government, imposes various duties on providers of online services with respect to harmful content, while at the same time protecting users' rights to freedom of expression and privacy (UK Parliament, 2021). The existence of hate crime laws represents "the official recognition of the harm of intergroup aggression and the importance of sanctioning it", as conducts motivated by a "performed negative opinion or attitude toward a group of persons based on their race, religion, ethnicity or national origin, or sexual orientation" (Boeckmann & Turpin-Petrosino, 2002, p. 208: Lawrence, 1999) have proven to cause a worse victimisation compared to ordinary crime, due to the targeting of a specific community based on immutable or prominent characteristics, which causes its members to feel unsafe (*ibidem*: Hamm, 1994). These data score the unique character of hate crime, which represents a form of victimisation to which mainstream institutions and culture are complicit (*ibidem*: Turpin-Petrosino, 1999). Notice, however, how most official definitions lack a gender-specific reference to hate speech addressed towards women, despite its high frequency.

However, there is no consensus, broadly speaking, about the definition of hate speech per se. Cohen-Almagor (2011), for one, defines it as "bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics" (Amaral & Simões, 2021), whereas Gelber and Stone define it as "an expressive conduct capable of, or intended to, instil or incite prejudice towards a person or group of people on a specific ground, including race, nationality, ethnicity, religion, sexuality, gender identity or gender" (Gelber & Stone, 2007). Hate speech is manifested verbally, non-verbally and symbolically, and it can be expressed in ambiguous and metaphorical terms, making it difficult to identify when it occurs through stereotyping that is considered to be socially acceptable (La Paz et al., 2020). In the debate between freedom of expression and hate speech regulation, it should not go unnoticed that hate speech constitutes in itself a violation of the freedom of expression of certain groups and individuals.

2.3 Hate speech: gender perspectives

Public discussion of hate speech towards women has been on the rise due to its peculiar characteristics. Directed at women for being women, it does not merely harm a single individual,

but women as a social group. Democracy-wise, this arguably represents a crisis, due to its consequent silencing of women and limiting of their democratic participation. Its impact is especially detectable online, due to communicative phenomena such as "the lowering of individuals' inhibitions and altered dynamics of interaction and decision-making in online spaces" (de Silva, 2021, p. 3). Recent accounts include offhand sexist remarks, so-called *revenge porn*, and invectives towards influential figures. Cyber-harassment of women includes "sustained and tactical campaigns of invective, image-based abuse, and other objectifying speech", engaged in *cyber-mobs* which happen to include multiple attackers at a time (*ibidem*). Among the factors worsening the possibility of identifying attackers and taking action against them are, undoubtedly, the structural anonymity and invisibility of users carrying out such attacks (*ibidem*). The consequent subordination and silencing are systemic because they accrue to women on the basis of their sex, which constitutes an axis of discrimination within patriarchal societies: as a consequence, it derives a certain authority from a pre-existing oppression (*ibidem*). Hence, sex-based vilification does not only constitute a discriminatory harm per se, but it also feeds into other forms of violence against women, who typically feel threatened and humiliated and adapt their behaviours accordingly, "by policing their identities, speech, and movements or by leaving online and offline spaces and disengaging from public life" (*ibidem*). The consequent marginalisation and devaluation of women's speech ends up excluding them from full and extensive democratic participation. The existence of a sex-based gap in anti-vilification laws, as already mentioned, further proves the constitution of a loophole whereby, in patriarchal societies, women are silenced to such an extent that their vilification drives, in part, its own acceptance into those societies, constraining women to the point that they are unable to publicly and fully protest its harms: the gap may indicate that their vilification is not only abundant, but normalised, at the same time overwhelming and invisible (*ibidem*). Hence, sex-based vilification is an issue for both public life and the entire democratic community.

The importance of the gender-based nature of vilification is made clear by studies concerned with the prescription and proscription operated by gender roles and stereotypes and the type of response deviations receive. In particular, "gender stereotypes are closely linked to traditional social roles and power inequalities […]; violations […] are met with various forms of punishment and devaluation, many of which appear to stem from their prescriptive quality" (Prentice & Carranza, 2002). While, in some cases, the deviation from stereotypes may be assimilated and go unnoticed,

in others, subjects are punished for their violation through negative evaluations, attempt to modify their behaviour, and social isolation (*ibidem*). Coincidentally, risky behaviours for women are those that are more in line with full democratic participation: women are expected to be discrete, silent, good-looking, and accommodating, whereas their willingness to be heard, stand their ground, and speak their own mind is frowned upon, especially when their stances differ from the status quo.

2.3.2. The digital dimension of gender-based vilification: perception and response

The line between hate speech and hate crime is a blurred one, as vilification encompasses all behaviours, from minor norm-violating to law-breaking acts. The deviance of online interaction, in particular, depends on users' "inherent belief system of what is socially acceptable and what is not" (Wilhelm & Joeckel, 2018, p. 382). This is articulated along the dimensions of moral orientation and moral identity, the hypothesis being that the higher the salience of individualising foundations, which include the dichotomy of harm/care and that of fairness/reciprocity, the higher the intention to flag hate speech; on the contrary, the higher the salience of binding foundations, including authority/respect and purity/sanctity, the lower is the intention to flag such comments, especially when they are aimed at women and sexual minorities (*ibidem*). Politics-wise, conservatives thus result to be more likely to advocate actions that perpetuate existing power structures, endorsing traditional gender roles and aggression towards individuals deviating from them (*ibidem*). However, controlling for political orientation, gender plays a role as well: women tend to score higher in harm/care and fairness/reciprocity, but also in purity/sanctity, whereas men exhibit a justice orientation (*ibidem*: Jaffee & Hyde, 2000). As a consequence, flagging behaviour is more in line with gender-related prescriptive stereotypes affecting women, due to their higher tendency to care for others and be sensitive to harmful content (Prentice & Carranza, 2002). Similarly, while men are shown to be more frequent commenters in online fora, women tend to be more discrete and polite, "often appear[ing] to be aligned even when they disagree with one another" (Wilhelm & Joeckel, 2018: Herring, 2003).

To this, it should be added that hate speech perception also depends on whether one belongs to one or more of the targeted minorities, aside from whether one agreed with the statement to evaluate (Wojatzki et al., 2018). In extreme cases of hate speech towards women, for instance, gender does not make a substantial difference in perception; however, it does in more nuanced

circumstances (*ibidem*). Problematically enough, however, there is to this date no accurate dataset serving as a golden standard for hate speech detection against women (*ibidem*).

One of the most accredited forms of hate speech is what author Mantilla defines as gender-trolling. The term *trolling* was first coined by Internet studies scholar Phillips, and it indicates a "disruptive conversation of an entire community by posting incendiary statements or stupid questions onto a discussion board […] for the troll's own amusement, or because [they] have a generally quarrelsome, abrasive personality" (Mantilla, 2013, p. 563: Phillips, 2012). Unsurprisingly, "the majority of trolls on the English-speaking web are […] white, male and somewhat privileged" (*ibidem*). In its gender-related declination, gender-trolling is not merely done "for the *lulz*" (Mantilla, 2013, p. 563), but it often expresses actual beliefs held by the trolls. This type of attack is often carried out through the coordinated participation of numerous people using gender-based insults, vicious language, and credible threats; and it often distinguishes itself for the unusual intensity, scope and longevity of the attacks, and a fierce reaction to women speaking out or responding to the hateful statements (Mantilla, 2013). The harassment is about patrolling gender boundaries to ensure that women are kept out of typically male-dominated arenas, such as political and public discourse, as it is shown by targets being mainly feminine and feminist subjects (*ibidem*).

2.3.3 Twitter: a case study

The generalised inability to prosecute hate speech is due to a bunch of legal and contextual factors, including complications arising from transnational jurisdiction, inadequate legislation or provision by providers and platforms, investigative bodies lacking the skills and resources to properly carry out the task, and the sheer amount of abusive online behaviour overwhelming the legal system (Hardaker & McGlashan, 2015). As a public-facing social network, Twitter is, among others, one of the platforms which mostly exposes users to behaviours that pose risks to their personal well-being, safety, and security: this is especially problematic due to the diffused anonymity of users and the possibility to re-blog and cite content, reaching a wider audience. It has, in fact, been proven by previous studies that "anonymity can foster a sense of impunity, loss of self-awareness, attitudinal polarisation, and a likelihood of acting upon normally inhibited impulses due to *deindividuation*": individuals may not be "seen or paid attention to", but are rather perceived by

single users as a homogeneous mass, weakening one's perception of the recipient's feelings and of their ability to respond and hold users accountable for their behaviour (*ivi*, p. 82).

These issues have not gone unnoticed: Take Back the Tech[1], for one, is a global campaign connecting the issue of violence against women and information and communication technology. In 2015, they published a report on corporate policies, and Twitter ran a pilot project for users to report sexist harassment and abuse, with the aim of collecting data to analyse and improve responses to harassment on the platform (Şahi et al., 2018).

However, things changed drastically after billionaire and entrepreneur Elon Musk acquired the platform on April 14th, 2022. A primary rationale for his purchase was, in his own words, the reduction in moderation of the platform, according to a "free speech absolutism" (Musk, 2022). Data analysis shows that the purchase was followed by an immediate spike in Twitter users using one or more of the most common hate terms in the period leading up to Musk taking over the company, with the previous 84 times per-hour 7-day average use of specific hate terms jumping to some 4,778 times; potential reach of this content was in excess of around 3 million (Benton et al., 2022).

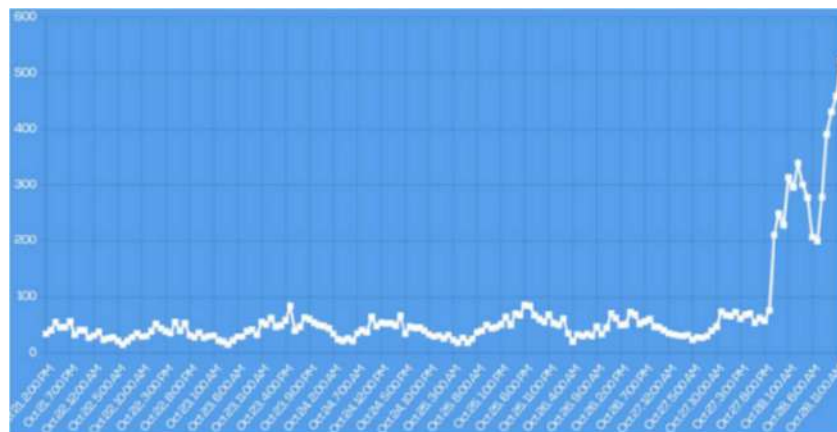Figure 1 - Spike in Tweets using one or more of the analysed hate terms in the period leading up to Musk taking over the company (Benton et al., 2022)

Sentiment analysis further confirmed that 67.2% of tweets using such terms expressed a negative tone, meaning that they mostly were not satire or re-appropriative interactions (*ibidem*).

---

[1] Take Back the Tech: https://www.takebackthetech.net/
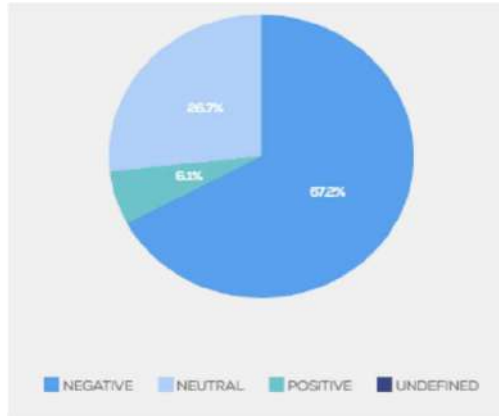
Figure 2 – Trendline on sentiment analysis spiking for all hate terms studied, with negative sentiment increasing almost immediately on October 28th (Benton et al., 2022)

Reductions in moderation after Musk's acquisition include the reinstatement of Donald Trump's account after it had been suspended for enticing the attacks on Capitol Hill on January 6th, 2021; the removal of the COVID-19 misinformation policy (Hickey et al., 2023: O'Sullivan, 2022); and the disbanding of the Trust and Safety Council, a board of organisations providing guidance on moderation (*ibidem.*: O'Brien & Ortutay, 2022). Unsurprisingly, previous findings on lowly moderated platforms, including Reddit and 4chan, indicate that such lighter moderation is associated with increased hate speech (*ibidem.*: Zannettou et al., 2018). Despite Musk's declarations, according to which he would ensure removal of bots and fake accounts on the platform, the latter actually increased after the purchase and spreading misinformation and hate (*ibidem.*: Uyheng & Carley, 2020).

Figure 3 - Overall hate speech on Twitter increased after Elon Musk bought the platform. Number of hateful tweets sampled during each week in 2022, compared to a baseline of tweets collected during the same time period. The spike in March 2022 coincides with the Canada convoy protests (Hickey et al., 2023)
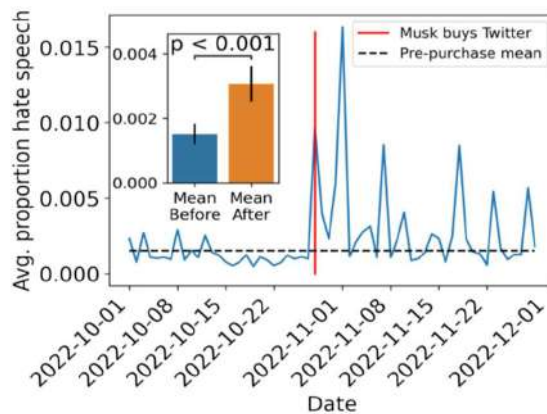
Changes in Twitter's content do not only threaten democratic debate within society, but also affect the company: advertisers, providing 90% of Twitter's revenue, have reduced their investments on the platform (*ibidem.*: Frenkel & Conger, 2022).

In response to inquiries, Musk has threatened to sue the independent researchers who documented the increase in hate speech, a response with no precedent from a tech company, potentially catalysing a chilling effect and frightening other researchers from studying the platform (Klepper, 2023).

2.5 Women in journalism and politics: a particularly vulnerable target

2.5.1 Online violence against women in journalism

A three-year global study on gender-based violence against women journalists online, covering 15 countries, comprehending 1,100 participants and examining 2.5 million social media posts has been conducted by the International Centre for Journalists and UNESCO in an attempt to analyse and address the "online-offline violence trajectory, […] representing a self-perpetuating cycle" whereby digital violence begets offline attacks and the latter in turn trigger online escalations (Posetti & Shabbir, 2022). The results are rather chilling: nearly three-quarters of respondents identifying as women declared they have experienced online violence throughout their work, with threats of physical violence rounding up to 25% of the total and additional threats towards people close to the subjects, including infants, amount to 13% (*ibidem*). In addition to gender-based vilification, racism, bigotry, antisemitism, ableism, homophobia and transphobia intersected, with black, indigenous, Jewish, Arab, Asian and lesbian participants recording the highest rate and most severe impact of online violence (*ibidem*). This type of violence comes in multiple forms, ranging from large-scale attacks and extreme threats to more networked and low-key, slow burning gaslighting (*ibidem*). 15% of participants reported experiences of image-based abuse, including deep fakes, and almost 50% reported being harassed through unwanted private messages (*ibidem*).

Unsurprisingly, online vilification is often accompanied by offline violence: one-fifth of respondents reported being attacked or abused offline in connection with online vilification, with mental health impacts being documented as the most frequent consequence (*ibidem*). In response to the attacks, 30% of respondents self-censored on social media, avoiding inflammatory topics such as politics and feminism; 20% withdrew from all online interaction; 11% were forced to miss work to recover; 4% quit their job; 2% abandoned journalism altogether (*ibidem*).

Among major triggers of online violence, gender has been identified by roughly half of the participants as the main story theme, followed by politics, elections, human rights and social policy (*ibidem*). 41% said they had been targeted by attacks flowing from orchestrated disinformation campaigns, and a surprisingly high number indicated political leaders and state actors as the instigators, scoring a terrifying 37% of the cases, second only to anonymous attackers in 57% of the instances (*ibidem*).

Tech companies are clearly identified as the main enablers of the attacks, hosting them and frequently failing to address or prosecute them: Facebook was the most frequently used platform, scoring 77% of the cases, closely followed by Twitter in 74% of them; however, the former registered a significantly higher number of incidents, occurring in 39% of the cases compared to Twitter's 26% (*ibidem*). According to the study, platforms not only are accountable for design and business model failures enabling such attacks, but they also lack comprehensive gender-sensitive and human-focused solutions, presenting inadequate representation in countries where women journalists are most at risk (*ibidem*). Transnational and cross-platform measures and responses are thus urgent, given that networked attacks often cross platforms and use languages which big techs lack the resources to easily identify and address (*ibidem*).

According to reports, journalists are the only category of public figures where women are disproportionately affected by abuse, receiving around three times the harassment of their male counterparts (Binns, 2017: Bartlett et al., 2014). In fact, while reported levels of abuse are similar between sexes, women are much more likely to receive sexual or profoundly threatening insults (*ibidem*). Women are overall more likely to report abuse across all categories, including their brand, their work and their personal sphere, in connection with their position as journalists (*ibidem*).

Some scholars consider abuse towards journalists to be a consequence of the openness of forums and comment sections, where readers enjoy the privilege of anonymity, hence why journalists' social media activities are considered to be a double-edged sword: on the one hand, it makes them more "down to earth" and thus more likeable, but on the other this professional informality exposes them to abuse (*ibidem*: Lee, 2015). As a consequence, many journalists struggle with their personal identity online, as they need to strike a constant balance with their audiences and sources (*ibidem*: Holton & Molyneux, 2015). In light of the optimistic idea that "a new level of responsiveness to the agenda of the audience is becoming built into the DNA of contemporary news work" (*ibidem*: Anderson, 2011), it is rather depressive to notice that the more comments a journalist receives, the more likely they are to see their audience negatively (*ibidem*: Meyer & Carey, 2014) and, in addition to the 16% of journalists who end up abandoning platforms altogether, 23% declare they changed their ways of researching and narrating within their job, indicating a chilling effect reinforced by the 15% that simply stopped following certain types of stories which would attract most abuse (*ibidem*). It most definitely does not help that the so-called glass ceiling is still firmly in place within the sector of journalism, with 3 out of 16 members being women in the Society of Editors and no woman being appointed to the Society's Advisory Council (*ibidem*).

Wider large-scale research, such as that conducted by the International Women's Media Foundation, provide additional and chilling data: around 65% of female journalists experience acts of intimidation, threats and abuse around the world, regardless of authoritarian or democratic governments (Sarikakis et al., 2021: IWMF, 2014). Worryingly and ironically at the same time, the abuse carried out under the guise of freedom of expression erodes that same right among women, as shown by a research from Amnesty International documenting the psychological impact of hate speech on women exercising their right to free speech on Twitter (*ibidem*: Amnesty International, 2018). Unsurprisingly, even in this scenario concerning specifically women journalists, hate crime perpetrators are twice as likely to be male, and the more serious the category of offence, the more likely are the perpetrators to be male (*ibidem*).

Further insight is much needed on the background of attackers. While the majority are isolated and anonymous, 16% are part of hate-oriented criminal gangs or groups (*ibidem*). This portion is possibly the most worrying, for it proves the acceptance of exclusionary and supremacist ideologies among certain strata, creating an often unquestioned "social reality" which can be normalised to the point where it is not adequately regulated (Tsesis, 2002). The issue is even more

severe once the role of leaders and political figures is added to the equation: it is not only a matter of traditionally accepted prejudices, but these are also reinforced by authoritative figures leading the way, reinforcing their followers' preconceptions and rationalising the maintenance of power at the hand of dominant groups (*ibidem*).

2.5.2 Online violence against women in politics

The idea of a double-edged sword applies to women in politics as well: while social media undoubtedly allows for a larger-scale political communication, encouraging the dialogue between politicians and users, it also exposes political figures to threats to their reputation, fostering political polarisation and possibly leading to personal safety risks for public individuals (Solovev & Pröllochs, 2022). Gender differences can be noticed in how communication in this field occurs: women are less likely to express negative sentiments and more likely to focus on moral dilemmas (*ibidem*: Nguyen et al., 2008), and hence are more likely to report messages targeting minorities; such differences are reinforced by traditional gender roles, according to which "women are perceived as less persuasive and are often outright dismissed when displaying aggressive behaviour online" (*ivi*, p. 2: Winkler et al., 2017). Together with other minorities, women tend to be a target of cyberbullying and attacks especially when they present strong, active and ideological stances such as feminism (Hardaker & McGlashan, 2016). Being critical more in general attracts a higher portion of hate, as shown by negative sentiments being answered with a sheer volume of hate speech especially towards left-wing individuals (Solovev & Pröllochs, 2022). Since such a response can hinder people to actively participate in politics (Scott, 2019), as well as increasing polarisation, "hyper-partisanship and less common ground between opposing political sides" (*ivi*, p. 4: Finkel et al., 2020), it can threaten the functioning of democracy itself by impeding a diverse composition of political institutions (Solovev & Pröllochs, 2022).

Many case studies have been conducted, especially throughout electoral campaigns. One such case concerned the so-called *chilling effect* on women in politics during the US 2016 campaign, which provided insightful observations. The main consequence of the chilling effect catalysed by hate speech towards women within the political sphere is to limit civic engagement online and prevent the fostering of digital citizenship (Citron & Norton, 2011). Additionally, it may discourage women from running for office in face of the harassment experienced by women candidates (Carlson, 2017). In the US, the protection of hate speech provided by the First Amendment

protecting freedom of expression has enormous social consequences, as members of defamed groups and minorities internalise the message of their inferiority (Matsuda, 1993). In the case of 2016, both women candidates and women reporters were targeted when carrying out or covering political campaigns, hindering the functioning of the free press as well (Carlson, 2017). Defined as a phenomenon occurring when "individuals seeking to engage in activity protected by the First Amendment are deterred from doing so by governmental regulation not specifically directed at that protected activity", the chilling effect indicates the vacuum generated by the gaps in governmental regulation concerning the specific right to freedom of expression of minorities and defamed groups and individuals (*ibidem*). As a consequence, in 2016, women from different political sides were dissuaded from activities characteristic of civic engagement, including posting on social media or speaking openly about their political preferences (*ibidem*). Although censorship is no longer inherent to state politics, then, the gap in regulation protecting minorities ends up privileging powerful groups and viewpoints (*ibidem*). It does not help that women are, from the start, underrepresented in the public sphere: in the US, women comprise 25% of sources and subjects in political and government-related news (GMMP, 2015), constituting the source of only 38% of the stories (*ibidem*) and producing only 37,7% of news reports at the top twenty outlets (Women's Media Centre, 2017). Although the issue of misogynistic hate speech has been acknowledged by the Supreme Court in some instances (RAV v St Paul, Minnesota, 1992), it should be a state goal in the US to establish the "essential pre-conditions for collective self-government", ensuring that all sides are presented and heard and that certain forms of protected free speech do not violate the equal right to it (Carlson, 2017: Fiss, 1996).

Despite these studies being conducted in the US within specific circumstances, they provide insightful information for other areas as well: Noelle-Neuman's *spiral of silence* theory, for instance, assumes that "individuals feel substantial pressure to conceal their views because they fear isolation, whereas those who believe their position as dominant will be more confident in public" (Noelle-Neuman, 1974). This is clearly applicable to multi-layered political positions: not only does it affect minority stances within any specific political context, but it furthermore affects minorities from multiple sides, as gender-based vilification, like any other form of abuse and discrimination, goes beyond the dominant or minority status of one's stance. As shown by studies conducted on the 2016 election campaign, all women from all political sides of the debate were more or less affected by the prevalence of online abuse directed towards them as such, with the

specific political context of the case making it all the worse, as it did not merely silence them as women, but it silence them as women *holding a political opinion*, *engaging in the civic and public sphere*, and *participating in digital democracy*. It is for these reasons that it is not enough to merely regulate hate speech as a form of vilification towards minorities, but the underlying democratic issues that it constitutes need to be addressed as a larger-scale social and civic problem: while the online domain and the technological sphere more in general hold undoubtful potential for democratic innovation, extended debate, higher inclusion and more effective engagement, the double-edged sword constituted by its lack of regulation – or the inabilities and limits to its application – constitute a major obstacle to its democratic functioning. Besides regulation of freedom of expression in accordance with principles of equality and the possibility for everyone to engage without their dignity and safety being at risk, it is of the utmost importance that democratic structures are implemented within the digital domain to actually bring out its potential for democratic innovation, fostering inclusive civic engagement and empowering politically disenfranchised groups and individuals.

2.6 Hate speech regulation and fostering of digital democracy: insights and potential resolutions

Compared to other types of crime and abuse, empirical evidence shows that the impact of hate crime victimisation is higher (Boeckmann & Turpin-Petrosino, 2002: Garcia & McDevitt, 1999), due to the fact that it transcends the individual's dimension: the realisation that one's community be targeted because of "immutable or prominent characteristics" slowly erodes feelings of safety and security (Boeckmann & Turpin-Petrosino, 2002: Hamm, 1994). Additionally, there is a historical continuity of hate crime victimisation, supported by the complicity of mainstream institutions and culture (Petrosino, 1999). While some people see this distinction as problematic, valuing one victim more than others, it actually underscores the unique qualities of hate crime (Boeckmann & Turpin-Petrosino, 2002). Again, then, hate crime configures itself as a multi-layered phenomenon, which does not only impact its direct victims but their community as a whole and, therefore, society at large. The silencing of prominent individuals from minority groups, as well as the groups themselves, triggers fundamental issues in the public sphere, which affect

disenfranchised individuals' and communities' ability to express themselves, participate in public engagement and, more generally, in the democratic process.

Current major arguments for hate crime legislation reflect on the difficulties of its enforcement (Boeckmann & Turpin-Petrosino, 2002: Gerstenfeld, 1992), the establishment of 'protective' status categories for its victims (Boeckmann & Turpin-Petrosino, 2002: Jacobs & Potter, 1998), the constitutionality of statutes dedicated to the issue (Boeckmann & Turpin-Petrosino, 2002: Lawrence, 1999), and the issue of the perceived devaluation of other victims (Boeckmann & Turpin-Petrosino, 2002: Grigg, 1992). Minorities protection counts many precedents, including laws ensuring their rights and their freedom from persecution; the application of such laws in appellate courts further proves the existence of a trend concerned with tolerance and protection of diversity, which finds its expression in law enforcement practices focused on monitoring bias-motivated crime (Boeckmann & Turpin-Petrosino, 2002).

The translation of this stance into actual politics presents, according to Boeckmann and Turpin-Petrosino, a threefold objective, developing on different levels of administration. In line with Fraser's reflections on subaltern counterpublics, community-based public fora need to be institutionalised in order to allow the public to engage in a dialogue concerning social values: society at large needs to be informed of the consequences hate speech has on targeted communities and, consequently, on the all-encompassing value of equality. As hate speech is inherent to existing social structures and hierarchies, it is fundamental that the public at large develops a certain degree of consciousness of its own position and role in the matter, embedding the prime relevant values of inclusion and equality in deliberation concerning hate crime laws (*ibidem*). Moreover, to ensure that the public is started off young on the issue, school-based curricula should be dedicated to value training and the harmful impact of hate crime: this underscores the importance of one's exercise of civic responsibility in a democratic society, underlining the importance of civil liberties and constitutional protections, including freedom of expression, as well as one's role in preventing such rights from enabling social harm (*ibidem*). Finally, hate crime statutes should be strengthened and enforced legally, incorporating the public interest and relying on precedent law: in the case of the US, precedent *Brown v Board of Education*, for one, enshrines the principle of equal citizenship, stating that "every individual is presumptively entitled to be treated by society as a respected, responsible, and participating member" (Boeckmann & Turpin-Petrosino, 2002: Lawrence, 1993).

The concrete application of these policies should lead to a concerted work between legislatures and investigative authorities, as the rather sophisticated and complex outreach of violent groups has far-reaching social, political and law enforcement implications, especially in cases of retaliatory hate crimes, where intergroup relations are strained and tensions are high (Boeckmann & Turpin-Petrosino, 2002). Working coalitions able to either prevent or get in touch and intervene within hateful groups should encompass law enforcement officials, public school administrators, parent-teacher organisations and other community interest groups, providing education about hate groups; disseminating the information to community youth, families and educators; providing inputs to local enforcement authorities, including prosecutors and judges; and "issuing periodic reports to the public describing the activities of hate groups, their infiltration among local youth, and the type of bias-motivated hate crimes committed in the area" (*ivi*, p. 216).

However, to successfully carry out such tasks, research is needed to understand and quantify the impact of hate speech and better inform the debate around it: this can be done by analysing victims' responses and the trauma they report (*ibidem*). Studies concerned with the unique consequences of hate crime lend credence to the view that it is a civil rights issue; moreover, research suggests a necessity for institutional and transnational reform to effectively address the issue (Herek et al., 2002). Undeniably, the unique offences constituted by hate crime need to be framed in the context that minorities are pre-emptively well aware of the extreme violence that they experience; based off this already problematic start, hate crimes can embody a specific type of terrorism, inciting terror and controlling as well as manipulating victims, indirectly restricting their access to public spaces, education and the job market (Boeckmann & Turpin-Petrosino, 2002). Due to the hardly enforceable nature of measures against hate speech, soft policies can be integrated, such as diverting offenders to non-court programs to encourage their dialogue with victims and the community, to restore dignity through apology and reparations (*ibidem*). This aim includes a better service to hate crime victims, as minorities are shown to underreport hate crime, or are further victimised by insensitive or hostile authorities and public disclosure of their experience by other individuals; hence, more functional protocols for hate crime reporting are necessary, together with mental health provisions to support them in the process (*ibidem*). Ironically, one helpful element is minority social identity: victims lacking strong identification are shown to be more at risk for psychological damage, including self-blame, the internalisation of negative emotions, and the failure to seek redress; identity also plays a political role, determining support for policies

protecting certain groups and minorities from hate crime and hate speech: hence, groups seeking to generate broad public support for policies against hate crime, should emphasised ascribed characteristics over achieved ones, which only concern primarily specific individuals (*ibidem*).

Online, hate is expressed multimodally, constituting complex forms of expression which make it hardly decipherable. The lack of appropriate data and robust management tools make it hardly assessable, especially as its dynamics vary across platforms, despite its ability to cross them and spread across different networks, through "interconnected hate clusters from global 'hate highways' which cross social media platforms, sometimes using 'backdoors' after being banned, as well as jumping between countries, continents, and languages" (Vidgen et al., 2021). Despite its immaterial nature, online hate speech has been repeatedly shown to be connected with offline forms of hate, constituting temporal and geospatial connections which are often built upon trigger events such as political elections and terroristic attacks (*ibidem*). The functioning of platforms themselves plays a huge role in creating *rabbit holes*, with algorithms promoting new content along the lines of previously visualised hateful content, and comment sections, just like any other fora, can easily turn into hateful spaces, open to malicious actors and groups (*ibidem*).

The main issue with regulating hate speech is and remains the lack of consensus around its definition, characteristics, and limitations, across national governments, intergovernmental organisations, Internet platforms and civil society organisations. Two main aspects, fundamental to the understanding and addressal of online hate, are determined by context: the latter affects whether content should be considered hateful, depending on the identity of the creator, who may be a target reclaiming the terms and slurs used against them; moreover, it determines the impact of online hate and the harm it causes, as public and visible figures may reach wider audiences (*ibidem*). Hence, both individuals on the production and receiving ends of potentially hateful content determine, through their subjectivity, whether such a status applies.

There are multiple ways in which hate can inflict harm, depending on content and context, including "immediate distress and emotional harm, long-term mental health effects, a long-term impact on the victim's behaviour", a negative impact on the victim's – and other disenfranchised members' – willingness to engage in public and civic forums, the motivation of offline attacks as well as other online attacks, and wider implications for social justice and fairness (*ivi*, p. 4). Moreover, online hate can be configured as threats, which express the intention to engage in

harmful actions; incitation, which encourages, advocates, or justifies harm; demonisation, which likely inspires hatred indirectly; and animosity, which expresses prejudices against a certain group, albeit indirectly (*ibidem*). As it is clear, some hate involves amplifying elements, which heighten the tension and vitriol, whereas, in other cases, content is more hardly identifiable, resulting in ambiguous statements sometimes working as *dog whistles*, when they are apparently innocuous but reference elements clearly harmful to the eyes and ears of the victims (*ibidem*). Two orthogonal elements which help identify hate and address its impact are *hazard* and *influence*: the former is the potential of content to inflict harm in a given context, depending on its features, its overtness, and its broader social, historical and political references; the latter represents the reach and resonance of such content. Along these two dimensions, four main quadrants can be drawn: dangerous speech is highly hazardous and has substantial influence, occurring when public figures use explicitly hateful language and elements; bedroom trolling is highly hazardous, but low in reach; benign viral content is either neutral or positive, and reaches many people; everyday talk is not hateful and registers a low influence (*ibidem*).

2.6.1 The means of intermediaries: content moderation systems, friction, and user involvement

Online hate is primarily addressed through content moderation systems, through which the policies and values of the platform shape the technology and process at its base. Given that large tech companies have a substantial infrastructure of people, process and technology to shape and implement such choices, failures often reflect the values and biases of the creators themselves. Although the characterisation of moderation systems depends on the platform and its functions, some desirable features are common to most: moderation systems need to be high-performing, fair, robust, explainable and scalable (*ibidem*). This entails that not only must they be precise and work equally across different groups, but their decision-making process must be made clear and understandable in order to account for their cost-effectiveness and their environmental and social impact (*ibidem*).

Most platforms, however, do not operate autonomously in hate speech detection, and while 'stacked' content moderation is useful in that it extends definitions and limitations to different platforms, it also monopolises the process at the hand of a reduced number of service providers (*ibidem*). The roles of content moderation include characterising online hate, namely defining it,

providing a framework for its classifications and guidelines to carry it out; deploying strategies that can be used to identify it; handle it through a proportionate response; and enabling users' complaints through accessible review procedures (*ibidem*). For the process to be as complete as can be, user reports should be reviewed by both AI and human moderators, although the latter's employment can be time-consuming, expensive and inflict significant damage on them, as they are constantly exposed to a huge mole of hateful content: many moderators are notoriously underpaid and receive inadequate care and support, with some embracing extreme stances and others making a significant use of substances to deal with their job's side-effects (*ibidem*). Due to its costs and externalities, moderation is often outsourced to lower-cost sites, which implies not only the exploitation of workers from developing countries, but also impairs the performance of the moderation itself, given that online hate is often nuanced and contextual and thus requires a range of cultural and social factors to be addressed according to the expectations (*ibidem*). The great promise of AI thus lies in reducing human moderation, increasing the speed at which hate is detected and minimising bias and inconsistencies through a training on large datasets, which subsequently enables AI to label new content; however, its limitations can be serious, depending on the content they are processing and how decipherable it is, for instance in the case of video processing (*ibidem*). Among the main limits of AI are a lack of understanding of the wider social and historical context, as well as the speaker's identity and their previous online activity; a struggle with understanding satire and other complex forms of expression; a high bias with different social groups; frequent failure to identify hateful content as part of a *conversational dynamic*; issues with identifying hate within multimodal content; and a failure to consistently decipher coded language (*ibidem*).

Among the strategies employed by service providers to tackle online hate is the imposition of a certain degree of friction, which determines the level of resistance a content encounters in order to be published or found by large audiences (*ibidem*). While banning users altogether represents the highest degree of friction, intermediary measures can be taken, including "hosting constraints, viewing constraints, searching constraints and engagement constraints" (*ivi*, p. 5). One instance of this strategy is that of *shadow-banning*, occurring when a user is not visible to the audience at large, but can still publish their content. However, this strategy also carries its limitations: limiting users' ability to post and reach a wide audience is a de facto restriction of freedom of expression, which not only reduces engagement, but often requires a sheer volume of data to track users'

behaviour (*ibidem*). Moreover, given that detection systems are imperfect, non-hateful content may be deemed hateful and restricted, generating a chilling effect around a certain issue or topic and thus penalising free discourse (*ibidem*).

In this landscape, it is of the utmost importance that users be provided the possibility to appeal to decisions related to moderation, given the various above-mentioned limitations of moderation systems. Three key areas can be identified where this is mostly relevant: in the field of the information users are given, in their involvement within the content moderation process, and in the speed at which content is moderated (*ibidem*). As already mentioned, restrictions are sometimes not applied to the publishing of the content itself, but rather to its reach: in the case of shadow-banning, users are usually not notified about the intervention, for it mainly affects research of their content from third sides. In other cases, usually when direct interventions are applied and content is taken down, users are usually informed. The involvement of users in the moderation process also carries a certain weight, given the sheer volume and the rapid change in forms of hateful content: the ability of users to not only flag and signal, but also be informed about decisions made about a certain type of content, creates a two-way street facilitating the entire process and its adaptation in time. At the moment, most platform policies do not include sustained community engagement, and merely allow little feedback from those affected by moderation decisions: opening up the Pandora box behind the entire moderation process to users represents not only a potentially more efficient way of handling hateful content, but may also foster dialogue and multilateral engagement, providing an opportunity for digital democracy and digital citizenship within platforms. The lack of consistent communication between service providers and users can in fact lead to the adoption of inadequate policies, or to a lack of speed in their update (*ibidem*). To bypass the issue, some companies are thus exploring new frontiers of content moderation: Facebook, for one, established an Oversight Board to help decide on "significant and difficult" issues in content moderation, constituting an entity composed of independent members who rule such cases (Facebook, 2018). Lastly, the speed at which content is moderated also constitutes a major venue for minimising the impact of online hate, in accordance with the EU Code of Conduct to Tackle Online Hate, which mandates platforms to review flagged content within twenty-four hours (EU Code of Conduct, 2016): this represents another area in which a higher involvement of users in the process would likely smooth the process.

2.6.2 Limits and issues of content moderation

Content moderation as a strategy to tackle online hate presents, regardless of its structure, a number of limitations which cannot be ignored. One of the main debates is represented by the constraints it imposes on freedom of expression, which need to be weighed against the harm caused by hateful content. As already mentioned, the removal of content which is non-hateful or expresses veiled and ambiguous forms of prejudice which do not, however, produce serious harm, is one of the risk of structurally imperfect moderation systems. The chilling effect generated by such removals could turn out to be counterproductive, undermining politically sensitive perspectives and important emancipatory activism. As a consequence, users may be inadvertently encouraged to migrate to smaller, less-regulated platforms, causing service providers to lose revenue and activities. Of course, as it has already been brought up, a lack of restriction of hate speech can constrain freedom of speech just as much if not more, preventing targeted and disenfranchised groups from feeling comfortable and safe in taking part to public discourse and democratic debate (Vidgen et al., 2021).

Another key concern is that of privacy, in conjunction with anonymity and, most importantly, data use: personally identifiable information can be used as an input into an AI system, regardless of whether it is stored with the user's explicit consent and whether it is visible to moderators. The sheer amount of data necessary to a correct functioning of most moderation systems not only carries significant costs, but studies also show that users value their personal data highly and seek to protect them at all times (*ibidem*).

2.6.3 Further roles of intermediaries in fostering digital citizenship

Aside from negative responsibilities such as content moderation, it is fundamental that intermediaries also recognise their positive responsibility in challenging digital hate by fostering digital citizenship, protecting user's right to take part in the Internet's diverse political, social, economic, and cultural opportunities, which inform and facilitate civic engagement. Among the means available for intermediaries is increased transparency, which requires to engage in conversations with internal and external stakeholders to identify the potential harms of hate speech; various actions can subsequently be taken, which "reflect the ways in which their different activities may facilitate the spread of online hate and thus undermine digital citizenship" (Citron & Norton, 2011, p. 1441). The role the Internet plays in facilitating civic engagement facilitates in

turn democratic functions, as citizens build networks of social interaction and trust, learning habits of cooperation and public spiritedness. Basically, the Internet constitutes a public space of engagement, information, dialogue and education, representing a potentially all-inclusive meta-forum acting as a mediating institution cultivating lines of trust across lines of social division (*ibidem*). On the other hand, however, it can also foster isolation and disengagement by distracting people from face-to-face interactions; plus, the possibility for an unregulated, horizontal, anonymous and pseudonyms discourse can easily accelerate destructive behaviour, with history and social science confirming that "hate speech may facilitate acts of violence against members of targeted groups", ending up with silencing them and inflicting serious psychological injury (*ivi*, p. 1448). Moreover, hate speech may further degrade public discourse by skewing society's assessment of members of certain groups of their ideas: by devaluing their expression, hate speech can produce a process defect in the marketplace of ideas. Furthermore, because hate speech may inspire or deepen prejudice, it can lead to discriminatory decisions about jobs, housing and life opportunities as stigma can render targeted groups dishonoured and erect significant barriers to full acceptance into the community, undermining social commitment to equality and dignity: search engines ensure the persistence of cyberhate and its costs to digital citizenship, as it produces "permanent disfigurement" of group members (*ibidem*). Due to concerns related to potential business, moral and instrumental costs of cyberhate, intermediaries are then beginning to tackle the issue in terms of commitment to digital citizenship; however, it is hard for them as private actors to define key terms such as "hateful speech" specifically, and although most commentators agree that its harm is sufficiently serious to warrant prohibition, consensus around its regulation and how it should be carried out is exceedingly rare (*ibidem*). One way intermediaries can circumscribe the definition of hate speech within their policies is to set the bar at what is considered to be actionable harassment in other contexts, such as school and work: elements determining the seriousness of hate speech in these areas include its frequency, severity, threat, humiliation, and offence inciting psychological harm (*ibidem*). Otherwise, intermediaries may define speech as that which silences or devalues its targets' counter-speech, shutting down reasoned discourse: this can be derived from private institutions' regulations, especially universities and other educational entities, which hold extensive experience in regulating this kind of speech in order to tell it apart from "rational but offensive propositions that can be disputed by argument and evidence" (Byrne, 1991).

2.6.4 Additional solutions to tackle online hate: media literacy and counter-speech

Given the complexity of the matter, tackling online hate requires a multi-level approach including all stakeholders in the digital domain. Transversal solutions range from the sphere of education to that of political regulation, to service providers' responsibility in handling the issue. The installation of hate speech policies "at respected and prestigious institutions has a 'trickle down' effect" (Sarikakis et al., 2021: Jon, 2005); however, it is equally important to note that although some online platforms can setup policies to address hate speech, they may be slow to implement them, as the provisions "are often broad and vague, causing confusion, but also leaving platforms the flexibility to use these policies to their own needs", and this shows that self-regulatory measures may not effectively address the issue of online abuse (Sarikakis et al., 2021). Due to the dichotomy between hate speech regulation and freedom of expression, most major institutions, including the European Union, favour 'soft' measures, drawing from grassroots movements and initiatives: #MediaAgainstHate, for one, is a Europe-wide campaign led by the European Federation of Journalists, which "provides support to journalists exposing hate speech who have become targets and victims of hatred and harassment for speaking out" (*ivi*, p. 71: EFJ, 2018). The No Hate Speech Movement, on the other hand, is a youth campaign of the Council of Europe for human rights online, aiming to "reduce the levels of acceptance of hate speech and develop online youth participation and citizenship, including internet governance processes" (*ibidem*: No Hate Speech Movement, 2018).

From the educational perspective, media literacy, defined as "the ability to use, understand and create media and communications in a variety of contexts" (Vidgen et al., 2021), has been proposed as a potential way of tackling the spread of harmful content online, increasing users' resilience and critical faculties and enabling them to counter and challenge not just hate but harmful content. Media literacy can take many forms, from providing users with training and information about being online, to changing the design and functionality of online platforms; it could thus have a positive effect in tackling other online harms as well, such as misinformation (*ibidem*). As for now, no large-scale studies actually establish whether media literacy works as a form of countering online hate; moreover, most forms of media literacy are only likely to address part of the problem, as individual-level interventions may "help targets of hate and users at risk of becoming hateful,

but are likely to do little to address the more committed and entrenched purveyors of hate" (*ivi*, p. 102).

Among educational efforts, intermediaries can play a significant role as well, helping to "develop an understanding that citizenship should include attention to the dignity and safety of others": they can "educate their users about digital citizenship norms by transparently explaining their enforcement", possibly drafting a set of principles and explanatory examples; the more they do this, the more informed users' choices are about the sort of online community with which they interact, as well as "engage in efforts to educate the public more broadly about hate, for instance by funding cyber literacy campaigns to teach students about digital citizenship" (Citron & Norton, 2011, p. 1478).

The major discussion is, however, that around counter-speech, which can be understood as "content which challenges, undermines, criticises or calls out hateful content: it has attracted support from advocates of free speech who view it as a way to contest and challenge hate without needing to constrain freedom of expression" (Vidgen et al., 2021, p. 101). A multi-faceted counter-speech is the ideal approach to counter hate speech, embracing a multiplicity of legal strategies which incorporate regulatory and non-regulatory counter-speech measures, which may come from targets themselves as well as from actors speaking on their behalf, as part of a collaborative effort to foster a holistic and effective process of cultural change (de Silva, 2021). Whether counter-speech is successful in changing the viewpoint of haters "will depend upon many factors, including how the counter-speaker engages with them, their own attributes and the outlook and proclivities of the hater"; however, the purpose is not to change the mindset of the hater, but to support the victim (Vidgen et al., 2021, p. 103). As with any intervention, counter-speech entails specific costs: the fact that "it is typically created by socially minded individuals or by community groups, many of whom have either experienced online hate or are actively involved in efforts to tackle it, inevitably places a burden on users who may have already been burdened by being targeted" (*ivi*, p. 103). Furthermore, the focus on counter-speech shifts attention away from structural factors enabling online hate "to be posted, shared and to reach large audiences instead of emphasising how individuals can address it; other concerns pertain to the risks inflicted on the counter-speakers themselves, as they may put themselves at risk of being targeted" (*ivi*, p. 104). One partial solution would be to resort to bots which automatically generate counter-speech, minimising the exposure of humans to hateful content and providing a large-scale counter-speech campaign that goes

beyond merely reaching out individually to potential haters; however, automatic and thus uncontrolled bots can not only accidentally push users to become more hateful or reduce users' trust in platform policies, but their liable design can also expose them to the risk of learning bad habits and being exploited by hateful users (*ibidem*).

Nonetheless, the potential of counter-speech lies in its potential involvement of additional actors, including intermediaries, which can counter digital hate with speech of their own. For now, this is extremely rare, which constitutes a significant problem given the costs that digital hate carries: "because intermediaries often enjoy respect and a sense of legitimacy, their counter-speech can demonstrate what it means to treat others with respect and dignity; moreover, such counter-speech can expose digital citizens to diverse views, piercing the insularity of hateful messages that may lead to more extreme views" (Citron & Norton, 2011, p. 1474). By using counter-speech, intermediaries can help transform online dialogue by documenting the existence of hatred while concomitantly rebutting it, which may help develop the qualities of tolerance and repair the public discourse by speaking for silenced or devalued targets (*ibidem*). Indirectly, intermediaries can sustain counter-speech by empowering users to respond to hate speech and report TOS violations, enforcing community norms and expectations of digital citizenship; however, such efforts are "more likely to be effective when intermediaries have educated their users and enforcement personnel about the specific harms to be addressed by their own hate speech policy" (*ivi*, p. 1478). In fact, communities that permit "mutually visible action among the participants, credible commitment to shared goals, and group members' ability to punish infractions" create contexts in which users "can do a better job in managing the resource and policing infractions that can markets or government systems designed to accomplish the same goals" (*ibidem*: Ostrom, 1990). In an increasingly democratic vein, "intermediaries can also help encourage the development of digital citizenship norms through architectural choices": the web's anonymity, for one, was neither an inevitable nor a necessarily healthy feature of web design, as the Internet's communicative strengths also magnify its capacity to empower socially destructive behaviours (*ivi*, p. 1482).

Among the various layers of counter-speech's functioning is the so-called *Butterfly Politics*, whereby small attempts to change inequality can have big impacts on the political system: these "include counter speech in the form of education, outreach efforts by governments and NGOs", and expanded hate speech politics in social media (Carlson, 2017: MacKinnon, 2017). In the US, several NGOs have been working to address the problem of hate speech through outreach and

education, and globally there exist many state-funded efforts (*ibidem*). In Europe, the Youth Department of the Council of Europe has funded a two-year campaign aimed at minimizing the use of hate speech worldwide: as a result, 45 countries inside and outside the EU have launched their anti-hate speech or anti-discrimination campaigns, especially among young people (*ibidem*).

From their part, users are collectively responsible for acknowledging and countering hateful content as part of their active, democratic duty as members of a community. Public opinion data suggests that most users wish to censor hate speech in the name of civility: 40% of millennials declared that they supported efforts to limit it and, in response to declining users, Twitter suspended the accounts of several high-profile members of the Alt-Right movement (*ibidem*). This responsibility is extendable to other components involved in the research and study of hate speech, but solidarity is often missing: when the interviewed women in media and politics "received inquiries from scholars who wanted to research hate speech, they said they felt valued and were committed to support the researchers, [but] by and large, the interviewees maintained that men did not know the feeling of receiving gender-specific hateful content as they were not aware of that issue and did not show sufficient empathy" (Sarikakis et al., 2021, p. 76). According to most interviewees, in fact, "men would listen, but they would not start a conversation about the matter: the men who asked and were supportive turned out to be close friends of the victims rather than colleagues from work, and men who downplayed hate speech tend to be in privileged positions and do not want to be confronted with the issue" (*ivi*, p. 77). In contrast, the female journalists felt more comfortable in seeking support from other women, including colleagues, friends, and women who also experienced hate speech or conducted research on the topic (*ibidem*). On online platforms, women can exchange their experiences with hate speech and abusers, discussing procedures and potential risks; in some instances, "they encouraged other women to block or report their perpetrators and, generally speaking, women declared their solidarity with each other and were more sympathetic, including female police officer" (*ivi*, p. 78). Even within their own workplace, targets report extreme differences in handling the situation: some workplaces offer support in providing access to therapists and treating the issue seriously, through small gestures and proactive attitude; in some other, however, the topic often remains ignored and not talked about (*ibidem*).

Despite progress made by many media employers over the past years, only 25% of the interviewees in the UNESCO-ICFJ research case said they had reported online violence to their employers:

while 10% gave no response at all, 9% advised the victims to grow a thicker skin and toughen up, with 14 respondents even being asked what they had done to provoke the attack (Posetti, 2022). This evidence of victim-blaming by employers is accompanied by the emergence of a pattern of restricting women journalists' speech as a way of managing online violence through social media policies introduced by news organisations: in some cases, women journalists reported being actively discouraged by their employers from speaking about their experiences of online violence or engaging with those attacking them; sometimes, they also got told to avoid discussing controversial topics on social media and, in the worst instances, there were cases of suspension or dismissal during a major attack (*ibidem*).

2.7 Concluding remarks

Drawing on academic literature, available data, and transnational regulation, this chapter attempted to shed some light on the current state of affairs around online abuse, particularly targeting minorities and especially those members that represent their disenfranchised group(s) in the public sphere. As we have seen, an intersectional approach is all the more necessary considering that the encounter of identities heightens the risk for abuse, especially within political debate. This is more so the case in Europe, where institutional and non-institutional violence towards minorities is higher than the global average, threatening the status of the European Union as one of the global bastions of democracy, pluralism and inclusion.

The apparent dichotomy between freedom of expression and protection from hate speech has recently been addressed by important public figures and organisations, including the United Nations (UN) Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human, and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. On, May 3rd, 2022, they issued a Joint Declaration on Freedom of Expression and Gender Justice officially stating that any limits imposed on offline and online hate speech in relation to gender- and sex-based discrimination are justified in the name of furthering gender equality, gender-based

inclusion, and the right of women to access the cyberspace and fully participate in online public debate (Joint Declaration on Freedom of Expression and Gender Justice, 2022).

Despite the Declarations' call for regulation, there remains significant gaps in addressing online abuse targeting women politicians, women journalists, and more generally women who carry a role in the public sphere. No measures currently exist covering this unique type of violence at the intersection of online gender-based abuse and violence against female public figures, which are not only targeted as women, but also as representatives, which makes the impact of such abuse on democratic quality all the worse.

As we have seen in the previous insights from digital constitutionalist literature usefully draw the full picture of the various public and private actors responsible for implementing an inclusive, pluralistic, and democratic cyberspace. This chapter touches on some of the main recommendations and possible routes for multilateral interventions, highlighting potential and shortcomings. Drawing on European measures, data, and interviews, the next and last chapter sketches the limits and merits of the European picture, building on what we know and calling for further research and work in the name of a truly pluralistic and gender-inclusive democracy.

3. CHAPTER: Online hatred towards women in the public sphere: a European case study

Despite there being a large corpus of studies developing in recent years around the topic of gender-based online harassment and its effects on public figures, most areas of the world still lack fundamental regulation addressing the issue. To be completely fair, public figures are not protected from other types of gender-based violence either, which constitutes a profoundly problematic loop, given that, as we have already seen, offline and online gender-based violence make up a destructive continuum of intimidation and humiliation, which often ends up victimising and delegitimising women's attempts at outgrowing imposed gender roles.

To this day, the only two countries in the entire world which have legally addressed the issue of violence against women in the political sphere are Mexico and Bolivia. In the latter, Law 243, passed in 2012, establishes prison sentences of two to eight years for different types of violence. To date, there has only been one successful sentencing for harassment and political violence, but efforts are underway, backed by national authorities, UN Women and other partners, to enforce the law (ACLED, 2021: UN Women, 2021). In April 2020, from its part, Mexico amended eight national laws to include provisions to prevent and prosecute violence against women in politics, as well as creating a National Registry of Sanctioned Persons for Violence against Women in Politics. In November 2020, it went as far as adopting guidelines to cancel candidacies for any person found guilty of committing gender-based violence. The National Electoral Institute has since received 36 reports of violence against women in politics (*ibidem*).

In comparison, European regulation is still far behind. The first to follow in the Latin American footsteps has been the Parliament of Catalonia, which, on December 28th, 2020, has passed Act 17/2020 that modifies Act 5/2008 on women's right to a life free of gender-based violence to incorporate, among other revisions, violence against women in politics. Pioneering this reform in Europe, the act defines violence against women in politics as 'the gender-based violence that occurs in spaces of public and political life, such as political institutions and public administrations, political parties, the mass media or social media' (Parliament of Catalonia, 2020). The act goes as far as recognising multiple forms of abuse, including physical, psychological, sexual, economic, digital, second order, and vicarious violence, which all constitute forms of

institutional violence (*ibidem*). This legislative reform implements the recommendations issued in recent years by the United Nations, the Organization of American States, the Council of Europe, the European Parliament and the Inter-Parliamentary Union (Verge, 2021). From their part, the United Nations have underlined multiple times how violence against women in politics goes against some of its milestones, including the Universal Declaration of Human Rights (1948), the International Covenant on Civil and Political Rights (1966), and the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW 1979) (United Nations, 2021); plus, the 2030 Agenda for Sustainable Development explicitly mentions as two of its objectives "full and effective political participation of women at all levels of decision-making" and the elimination of "all forms of violence against all women and girls in public and private spheres" (United Nations, 2015). Specifically focusing on violence against women in the public sphere, the United Nations' General Assembly's Resolution 73/148 states a deep concern "about all acts of violence, including sexual harassment, against women and girls involved in political and public life, including women in leadership positions, journalists and other media workers and human rights defenders" (United Nations, 2018). In that same occasion, the Special Rapporteur on violence against women outlined some explicit recommendations for national and transnational legislative authorities, including "[the] strengthen[ing] of complaint mechanisms and response protocols in line with international and regional standards by, for instance, issuing guidelines, codes of conduct and protocols for institutions like parliaments, electoral management bodies, political parties, electoral courts, legislative chambers or local administrations, and ensure enforcement mechanisms are functional" (United Nations, 2018). Compared to Europe, the Organization of American States has also been ahead in terms of raising awareness and preventing violence against women in politics, with a Declaration on Political Harassment and Violence Against Women being issued in 2015, in the attempt to "highlight the multiple manifestations of harassment and violence faced by women in the political sphere and foster actions to prevent, respond to and punish these crimes" (Organization of American States, 2015). In 2017, an additional Inter-American Model Law on the Prevention, Punishment and Eradication of Violence Against Women in Political Life was issues to codify not only the political rights that need to be acknowledged for women to access full political participation in the public sphere, but also the necessity for specific provisions to be adopted in order to enforce them (Organization of American States, 2017). Most recommendations

are similar in their very nature, underlying how institutions-led support at both the voter, the elective, and the institutional level is needed to achieve a full gender-inclusive political domain.

Studies on both European institutions as well as national, regional and local ones within European countries underline a pervasive presence of multiple forms of abuse towards women involved at various levels of the political domain. In comparative studies, some forms of abuse are prevalent in Europe compared to other areas of the world, such as mob violence (ACLED, 2021). Overall, women politicians face disproportionate risk in Europe, where violence towards them touches a peak of 26%, considerably higher than the global average rate of 12% (*ibidem*).

In order to better understand the depth and impact of this phenomenon, it may help to take a step back and pay attention to the composition of European institutions, as gender imbalances is the main factor laying the basis for abuse: the latter is nothing more and nothing less than a means of re-establishing gender roles, keeping women 'in their place' and reinforcing the idea that they are not welcome in the public sphere, especially in positions of leadership (OSCE, 2021). The urgency of the matter is all the more evident in light of the upcoming European elections, which follow the path of the previous terms, for better and for worse: while the absolute number of women in European institutions is increasing, their effective power is still limited to soft policies and secondary positions. The rate of violence they experience is still registered as high, both online and offline, both towards elected women and candidates, according to the most recent report, issued in February 2024.

3.1 The European landscape

Overall, most States in the European Union recognise freedom of expression as a fundamental human right, enshrined in both national and transnational law (Casarosa, 2020). Even before it was officially acknowledged as such by the Lisbon Treaty and the Charter of Fundamental Rights of the European Union, which integrated it as a right within EU primary law, the Court of Justice of the European Union had recognised its fundamental status within the European legal order (*ibidem*). Currently, then, the issue does not lie in its acknowledgment, but rather in its constitutional guarantee within national legal orders and its balancing with other fundamental rights (*ibidem*).

Among the norms, article 11 of the EU Charter grants a clear right to freedom of expression, by stating that:

"(1) Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.

(2) The freedom and pluralism of the media shall be respected." (Charter of Fundamental Rights, 2009)

Similarly, Article 10 of the European Convention of Human Rights confers a right to express and hold opinions, ideas and information without suffering from interference with authorities, specifying a right for the public to receive such ideas (European Convention of Human Rights, 1953). Aside from the negative obligation to abstain from interference, Member States may have positive obligations to protect this right even against private interference (Casarosa, 2020). According to the Article, interference with freedom of expression may be legitimate on three grounds: it must be protected by law, it must pursue a legitimate aim, and it must be necessary in a democratic society, corresponding to a "pressing social need" (European Convention of Human Rights, 1953). More specifically, the interference must pass a proportionality test, to ensure the possibility to adopt less intrusive measures: for instance, when freedom of expression conflicts with an equally protected fundamental right, and the consequent overriding requirement may be of public interest or lie in the protection of others' rights (Casarosa, 2020). This is because freedom of expression is not considered to be an *absolute right*, nor is there a hierarchy, within the European Convention of Human Rights, between *relative* rights (*ibidem*). The *margin of appreciation* is stronger in areas involving controversial moral choices, whereas it is weaker in others which entail fundamental rights of citizens, including political speech and criticism of the judiciary (*ibidem*). Aside from the clause on "respect for the reputation and rights of others", Article 10(2) lists a number of exceptional circumstances justifying limitations to freedom of expression, including national security, public safety, protection of health and morals, and so on (European Convention of Human Rights, 1953).

Nowadays, freedom of expression is also promoted within secondary law, through recent directives, Council decisions, and resolutions on specific matters concerned with the media (Casarosa, 2020). Among these, one of the most important regulations is the so-called General Data Protection Regulation no. 2016/679, which states in paragraph 153 the reconciliation between

the protection of data and freedom of expression as a responsibility of the Member States (General Data Protection Regulation, 2016). Moreover, Council Decision 2006/515/EC promoting cultural diversity and expression recalls human rights and promotes a diversity of cultural expression (Council Decision 2006/515/EC, 2006).

The same necessity to balance freedom of expression with other liberties is acknowledged in various national constitutions of Member States, some with an *ad hoc* limitation clause and others within a general one, investing domestic courts with the duty to strike the balance (Casarosa, 2020). Among the fields where conflicts are more likely to arise are reputation and honour, private and family life, dignity and image, and especially defamation and privacy (*ibidem*).

Among the limitations applicable to freedom of expression according to the European Convention of Human Rights is "expressions which spread, incite, promote or justify hatred based on intolerance", what has commonly come to be known as hate speech (European Convention of Human Rights, 1953). In the EU legal context, the most relevant provisions in this respect are the ones embedded in the Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law (Council Framework Decision 2008/913/JHA, 2008). As one can notice, other motivations, including gender, are not explicitly mentioned. Additionally, not all countries have appropriately implemented the Convention, and those that did followed different approaches, either including the offence in the criminal code or adopting special legislation on the issue, which may apply provisions different than those contained in the criminal code (Casarosa, 2020). Other EU legal instruments tackle the issue of hate speech in specific areas, such as the Audiovisual media service directive and the e-Commerce directive; interestingly, these include sex as a ground for protection (*ibidem*).

Recently, the approach of EU institutions regarding hate speech has moved towards soft law, namely forms of co-regulation whereby rules are negotiated between the Commission and private companies (*ibidem*). The Code of conduct on countering illegal hate speech online has been signed by the biggest private actors, including Facebook, Microsoft, Google, and Twitter, and it binds them to adapt their internal procedures by acting within 24 hours upon hateful content and notifying their users (Code of conduct on countering illegal hate speech online, 2016). The main issue with the Code of conduct is that it lacks a comprehensive definition of hate speech, and private companies applying their own conception of it may cause discrepancies with both EU and

national expectations (Casarosa, 2020). Most IT companies include in their codes gender, sexual orientation, disability or disease, age, etc. among the bases for discrimination, enforcing a higher level of protection, but privatising its enforcement at the same time (*ibidem*). Appeal mechanisms and their functioning are also left to private companies (*ibidem*).

3.2 The root causes: gender inequality in European institutions

Number-wise, European institutions hold a record of improvement in terms of gender-related inclusion in recent years. This is especially true for the European Parliament, which features as a prominent actor in the development of gender equality within the European Union, recording a higher number of female parliamentarians in comparison to the national average – what has been defined the *critical mass effect* – as well as introducing innovative ideas and actions promoting the gender cause on the part of these same female parliamentarians – denominated the *critical act effect* (Abels, 2015: Mushaben, 1998). This slight differentiation poses the accent on the difference between descriptive and substantive representation, the former being mainly formal, and the latter taking place when the involved subject effectively further their inclusion and promote their group's goals within the institutions they occupy (*ibidem*).

For these reasons, the European Parliament is, since the elections of 2014, "heralded as one of the most gender-equal elected bodies in the world" (*ibidem*: Fortin-Rittberger, 2014). Although the share of 25.8% of female MPs in 2014 was still far from real parity, the ever-growing percentage already represented a fundamental score in the face of the number of national female MPs decreasing in half of the European member states due to the rise of right-winged and far-right parties, but also due to a high number of parties gaining only few seats (often less than 5) and additional socio-economic factors (*ibidem*). Notably, left-wing parties recruit more women (*ibidem*: Fortin-Rittberger, 2014) due to their stronger adherence to the equality norm and their responsiveness to expectations of their constituents (*ibidem*). Overall, the European Parliament has been moving to the right, with the conservative European People's Party (EPP) being the largest political group, followed by the Group of the Progressive Alliance of Socialists and Democrats (S&D), the Tory-led group of the European Conservatives and Reformists, the liberal group ALDE, and the green and leftist parties (*ibidem*). In 2014, the share of female MEPs in the eight political group ranged from 21.9% (ECR) to 52% (GUE/NGL): in all political groups but the

ECR the number of women was above the average share of national parliaments, including the right-winged and even the radical right, with the Front National almost reaching gender parity as a result of the strict laws on *parité* of the French national system (*ibidem*). Moreover, in all leadership categories but presidencies of political groups, the number of women was even higher than the average of the European Parliament, with the number of female committee chairs rising to 10 out of 24. Yet, most prestigious and powerful committees were still chaired by male MEPs, "suggesting a disparity in influence between the genders" (*ibidem*: Renman & Conroy, 2014), with the more progressive leftist political groups no more inclined than the centre-right parties to nominate women as committee chairs (*ibidem*). Nevertheless, only 9 committees adequately mirrored the overall gender composition of the European Parliament, with the FEMM committee for Women's Rights and Gender Equality unsurprisingly registering as the most female committee, with only 10 out of 67 members being male. Generally speaking, while male-dominated committees are in charge of hard policies, the more female ones are responsible for soft issues – a pattern which can be found at the national level in most Member States as well (*ibidem*).

The European Commission is a different story: despite Juncker campaigning under the motto of "10 or more" in 2014, the Commission came to include 9 female members, which however keeps the record of one-third female commissioners going (EurActiv, 2014). This is especially positive in light of the fact that, for more than three decades, the Commission was an all-male body, improving throughout the 1990s. To promote this balance, Juncker promised the member states he would reward them with important portfolios in exchange for female candidates. In 2014, two out of seven vice-presidents were women (*ibidem*).

In celebrating the positive record of gender equality within European institutions, we should keep in mind that the so-called 'politics of presence' (Abels, 2015: Phillips, 1995) is not a sufficient condition to ensure effective equality of opportunities and power, as "political representation is best conceptualised as an active, multifaceted, and contingent process, driven by a broad swathe of actors with various views on group issues and interests, rather than as an authentic reflection of the values and needs of society by legislators though the vehicle of public policy" (*ibidem*: Celis et al., 2014). The FEMM committee, for one, is neutral and non-obligatory, a characterisation affecting its composition and its reputation, as well as its involvement: in fact, it is very rarely in charge of legislative proposals, acting as a consultative committee most of the time, which the MEPs regard as less relevant and powerful in comparison with 'hard policy' committees (*ibidem*).

While the programmatic homogeneity and the feeling of belonging stemming from this peculiar position may have some advantages (*ibidem*: Ahrens, 2012), the composition of the FEMM in recent years makes it rather problematic, as right-wing parties often exploit their participation within the committee to further a very conservative agenda when it comes to gender issues, contesting, for instance, reproductive rights (*ibidem*).

In terms of initiatives related to gender equality, the European Parliament was still, for the 2014 term, the main pioneer, adopting a resolution on June 9th, 2015, on "The European Union Strategy for equality between women and men post 2015" (2014/2015(INI)), which calls "on the Commission to draw up and adopt a new separate strategy for Women's Rights and Gender Equality in Europe, aimed at creating equal opportunities and based on the priority areas of the previous strategy with a view to ending all forms of discrimination suffered by women" (European Parliament, 2015). This resolution addresses a number of gender equality issues and states that strategy "must thoroughly take into account the multiple and intersectional forms of discrimination" (*ibidem*).

However, the self-image that European Union institutions portray as one of the most progressive gender regimes runs the risk of constituting an incontestable "myth" (Abels, 2015: MacRae, 2010) that obscures and leaves the more practical issue untouched: from a gender perspective, the 2014 was not that different from previous ones, embodying a certain reluctance to resort to hard law to address gender equality (*ibidem*). While the historical role of the European Parliament in developing the existing anti-discrimination regime is undeniable (*ibidem*: Kantola & Nousiainen, 2012), the perspective of intersectionality and the status of gender-based discrimination in comparison with other forms of discrimination is still unclear (Abels, 2015). The persistent involvement of women in the European Union constitutes a positive descriptive factor, but for its substantive representation to be efficient, a stronger linkage to transnational gendered publics is strictly necessary (*ibidem*).

In representative terms, however, the percentage of female Members of the European Parliament rose continuously after each election, reaching 41% following the 2019 election. This figure now stands at 39.4%, above both the world average for national parliaments (26.5%) and the European average (31%) (*ibidem*: Chiva, 2019). One puzzling result of the 2019 European elections compared to previous terms is that the proportion of MEPs from populist radical-right parties

increased, in some cases, quite substantially, mainly due to quotas, implemented transversally by political parties running for elections in countries like Italy, France, and Slovenia (*ibidem*). Within the European Parliament, women currently chair seven of the twenty-three standing committees and subcommittees, including budgetary control, internal market, economic and monetary affairs, and security and defence – a clear improvement compared to the previous 2014 term (European Parliament, 2023).

While the composition of the European Commission tends to fluctuate over time, there has been an improvement compared to the 2014 elections: after the 2019 term, one of the three executive vice-presidents is a woman, together with two of the other five vice-presidents (*ibidem*). Currently, the female commissioners hold a series of portfolios, including some major ones concerning the digital age, values and transparency, democracy and demography, health, transport, energy, innovation, research, culture, education, youth, and equality (*ibidem*). The president of the Commission in charge, Ursula von der Leyen, has committed to achieving full equality within the organ management by the end of 2024 (*ibidem*).

Additionally, in all European Union Member State national parliaments men outnumber women by large, and the situation at regional and local levels shows similar proportions, with an average of 32.1% of women in regional assemblies in the European Union (*ibidem*). The countries that classify as high achievers at national level are also the high achievers at regional level, and local and municipal councils are short of women representatives as well, with the EU average equalling 34.1% of women (*ibidem*).

Analysis of EIGE data found significant gender differences in the portfolios held by senior ministers in national governments: men are most often assigned high profile portfolios such as foreign and internal affairs, defence and justice, and finance and industry, whereas women are more likely to be given socio-cultural portfolios such as health, education or social affairs (*ibidem*).

3.2.1 Inequality in national institutions and civic engagement in Europe

Voter attitudes towards women are a key determinant of how many women are elected: while a very high percentage of responders in the EU declared that a female politician can potentially represent their interests, views on the existing distribution of power and women's role as leaders in decision-making positions are more nuanced. Among women, 62% think that there should be

more women in political decision-making positions in their country, while only 44% of men agree (*ibidem*). On average, 75% of respondents across the G20 considered men and women equally suited for government and politics, but not necessarily for top political posts (*ibidem*).

From a democratic perspective, which dives deeper into different strata of the political sphere, voter attitudes also matter in terms of civic engagement: many studies have shown that women on average tend to know or claim to know less about politics than men, a key issue for democratic participation (*ibidem*). No earlier than the 2019 EU elections, women were more prone than men to state that they did not vote because they did not know enough and less prone than men to state that they voted because they were interested in European affairs (*ibidem*). Thankfully, when it comes to actually voting, the gap between men and women decreases, and in the 2019 European Parliament elections, both women and men were more likely to vote than in the 2014 European elections, although the latter were still slightly more likely to vote (*ibidem*).

The so-called *democratic deficit* generated by gendered low participation, interest, and understanding of European Union affairs cannot be stressed enough, and European Union institutions are responsible for promoting not only gender equality within their own rows, but also in furthering gender-inclusive civic engagement. Recommendation 1676 on women's participation in elections, for one, "includes encouragement to parties to adopt quotas for lists, the use of 'zipped' lists, and training for women candidates" (Banducci, 2014, p. 3). One major issue playing a significant role within the gender-related democratic deficit within the European Union is the existing gender gap in public support for the institutions themselves: a lower level of support, trust, of knowledge about the European Union among women has been confirmed in recent Eurobarometer data, with almost half the women feeling that their interests are not adequately represented in the European Union, especially in post-communist countries (Renman & Conroy, 2014). Unless the European Union actively focuses on a higher involvement and consideration of women and their interests within its policies, it is likely that their voting attitude remains a significant barrier to their promotion and representation, in a loophole whereby women do not feel represented, do not elect women representatives, and do not see their interests integrated in European policies.

While research initially attributed women's political under-representation to a lack of political ambition on their part, studies confirm that structural and societal barriers continue to hinder

women (European Parliament, 2023, p. 7). Although these are not identical across the European Union, three interrelated factors emerge from general analyses. Barriers that hinder the 'supply' of women in politics include women's perception of their abilities and the costs of running for office, but also their unequal access to key resources, such as money, time, and political networks (*ibidem*). Unequal family responsibilities and a male-dominated political culture also influences women's willingness to put themselves out there, as traditional gender roles prescribe the private sphere as their social milieu. When women do find the courage to enter the public sphere, they continue to elicit strong, negative reactions, as shown by "the level of abuse and violence directed at women in politics and public life because they are women, and sometimes because of their increased presence" (*ivi*, p. 7). Additionally, limited 'demand' for female candidates reside in the way political parties themselves recruit, select and champion candidates, as well as voter attitudes towards women at different levels (*ibidem*). Finally, it has been established that "systems that are based entirely on proportional representation are more effective in promoting the election of female candidates, making certain electoral systems themselves a potential institutional barrier" (*ibidem*).

Among other institutions, the EIGE, the expert committee that advises on gender equality policy at European Union level, the European Women's Lobby (EWL), the Council of Europe Parliamentary Assembly, and research for the European Parliament promote a holistic approach and have identified a number of good practices to promote a wider inclusion in representative institutions, "ranging from 'hard' measures such as changing electoral systems and requiring legislated gender quotas, to 'soft' ones such as introducing voluntary party quotas and providing training, mentoring, funding and other support for women candidates" (*ivi*, p. 8). Findings on the impact of quotas were nuanced, but they generally appeared to have had a positive effect on women's chances of being elected (*ibidem*). The European Electoral Act could also require quotas for European Parliament elections, a possibility called for by Parliament, but rejected by the Council and not taken into account when the Act was reformed in 2018 (*ibidem*).

3.2.2 Inequality in media representation of women politicians in Europe

From the perspective of digital communication, research on the influence of media coverage of women candidates and politicians generally shows that under- and misrepresentation of women in media has a negative impact on women's aspirations and electoral success, although a higher media

visibility can help to get more women elected (*ibidem*). In fact, not all the attention brought to women in the public sphere is constructive: women are four times more likely to have references made to their appearance, and less likely to be challenged on issues (Lühiste & Banducci, 2016). Such biassed portrayals of women in the media also stifle ambition among women who would be willing to stand as political candidates in a better media environment (*ibidem*).

Compared to traditional media, social media have been and continue to remain a double-edged sword: women candidates' and politicians' use of social media to their advantage also increases their visibility in traditional media, as well as giving them a direct channel to reach the public and avoid mediated and potentially biassed coverage. However, as we have seen, social media also represent the number one place in which violence is perpetrated against women politicians, with 60.5% of the respondents declaring that intimidatory acts are "strongly motivated by the clear-cut positions taken on particular issues" and that taking a stance on "controversial subjects, such as defending women's rights and human rights" had made them particular targets of sexist behaviour and acts of violence (*ivi*, p. 10).

3.3 The consequences: violence towards women representatives in the European Union

Several Europe-based studies have shed light on the abuse and harassment women representatives generally experience throughout their career, although the entire domain of online harassment is still mostly untouched by research. A major one, carried out by the Inter-Parliamentary Union, involved 81 female members of parliament from 40 different European countries. Of these, 67.9% declared they had suffered sexist or sexual remarks in multiple instances throughout the course of their term in office, of which 35.6% occurred within the institutions themselves, and 22.1% on social media (Inter Parliamentary Union, 2018). Additional experiences of abuse also took place during political meetings, on the telephone or via email, in newspapers, on television and in their private lives (*ibidem*). The offensive remarks came from both colleagues from their own political parties as well as from others, plus voters and citizens (*ibidem*). Inappropriate compliments, disrespectful jokes and misogynistic observations made the subjects feel subject to personal scrutiny based on their appearance, as well as gender role stereotypes targeting their position in the public sphere (*ibidem*).

39.5% of the respondents expressed that newspapers and television broadcast images or comments about them which were highly derogatory or had a sexual connotation (*ibidem*). Moreover, 58.2% of respondents witnessed disparaged or sexual images and comments about them on social media (*ibidem*). Most respondents reported "being the target of repeated misogynous online insults and incitement to hatred, in which they were violently attacked as individuals and as women, particularly because of their political views" (*ivi*, p. 6). In some cases, the attacks included nude photomontages together with crude insults, as well as pornographic videos. 46.9% reported receiving "messages of death threats, threats of rape and beatings, against them, their children and their families" (*ivi*, p. 7).

Within their own rows, 24.7% of the MPs reported having been "sexually harassed during their term of office and 6.2% said that they had been sexually assaulted" (*ibidem*). In 75.9% of cases, "the sexual violence was committed by male colleagues, both from their own political party and from parties opposed to their own": 11.1% said that they had been "slapped, pushed, hit or had something thrown at them", and 6.2% said that "someone had threatened to use a weapon against them" (*ibidem*). Of the 40.5% of respondents that said that they had been "sexually harassed in the course of their work in parliament", and most cited "unwanted touching, inappropriate and persistent sexual advances and sexual blackmail" as the main case (*ivi*, p. 8). By including sexual comments into account, we end up with 1 in 2 respondents had been confronted with acts of sexual harassment at work. Financial abuse constitutes another slippery slope, with 7.1% of respondents being refused funds to which they were entitled (*ibidem*).

Out of these concerning numbers, only 53.3% of women parliamentarians who experienced abuse reported the incident to third authorities: this was the case for only 50% of those being directly threatened with death, rape, or beatings reporting it to the police, and 32.6% of those harassed online reporting it to technical intermediaries or other services (*ibidem*). Even so, most perpetrators were not held responsible for their acts, and participants underlined the fundamental lack of a service or mechanism within their parliament to care for such phenomena (*ibidem*). Where some services are in place, many doubted their effectiveness or fairness, underlining the tendency of their entourage to blame them for the event, insinuating that they were either lying about it or that they were the ones provoking it in the first place (*ibidem*). According to most women parliamentarians, what they experience is regarded as a natural part of their duty, despite it leaving

them psychologically scarred, with one-third fearing for their safety as well as that of those close to them (*ibidem*). Additionally, 33.3% stated that such acts had affected their freedom of expression and scope of action, admitting that they had become "more guarded" in their positions and statements, generally acting "more cautious and watchful" to become "a little less visible" (*ivi*, p. 10).

This phenomenon is "undoubtedly damaging to the reputation and image of parliamentarians and of parliament as an institution, its importance among other state institutions and how seriously its role and members can be taken" (*ivi*, p. 11). It also represents "a reflection of the violence perpetrated against women in European countries in general, where a whole host of political traditions and practices – drawn from cultures that are still largely patriarchal and male-dominated – are in use" (*ibidem*). Some respondents said that they "had been intimidated and attacked when seeking leadership positions, for the notion that a woman has made her way in politics by granting sexual favours is still very much alive", and the party loyalty required from politicians acts as a brake on disclosure and complaints: consequently, women fear of being seen as politically disloyal or weak if they speak up (*ivi*, p. 12). Among the reasons why such acts are so common in the political domain is a shared perception on the part of perpetrators that they hold exceptional power bestowed by parliamentary status, which grants them a certain degree of immunity and impunity (*ibidem*).

As part of the loop of violence, "women leading the fight against gender inequality and violence against women or speaking out on other sensitive political issues are often prime targets for organized attacks, including by ultra-conservative or anti-feminist groups", constituting further evidence that the main aim of intimidatory acts is to silence those women that challenge patriarchal norms (*ivi*, p. 13). To add to this, being young and belonging to the opposition have also been proved to be aggravating factors (*ibidem*).

3.3.1 The online dimension of violence towards women politicians in the European Union

"[T]he internet is not creating new forms of crimes against women […], but it is creating new ways and means for crimes to be perpetrated" (Bardall, 2013: Banks, 2001). The Internet, social media and more generally ICTs "have proven to be uniquely dangerous instruments in perpetrating

violence against women in public life, because of the relative importance of psychological violence in their political experience" (*ivi*, p. 2). Recent innovative and sophisticated forms of attacks on women include the use of 'spy software'; hacking; saved 'cookies' and browser histories; email tampering and interception; visual surveillance and geographic tracking via Global Positioning System software (Southworth et al., 2017). A number of the specific qualities of social media make them peculiarly suited to inflicting psychological violence on women in public life, due to women's unique vulnerability to attacks on the basis of morality: by breaking into a traditionally masculine field, women are exposed to sexualized and morally degrading criticism. The nature of messaging in social media facilitates ridicule, shaming and other psychological forms of violence against women in public life (*ibidem*). More generally, within public life, "women are increasingly becoming victims of violence as they join social movements and non-governmental advocacy networks to voice their political concerns, including defence of human rights" (*ivi*, p. 17).

Female MPs in Europe are particular targets of online attacks. 58.2% of those interviewed by the Inter-Parliamentary Union had experience of abusive, sexual or violent content and behaviour on social networks (Inter-Parliamentary Union, 2018). In fact, electronic communication is the primary means used for threats against female MPs. In 75.5% of cases, the perpetrators of these threats are anonymous citizens (*ibidem*).

Unluckily, despite the concerning numbers, comprehensive research on specifically online abuse towards women in the European political and public sphere does not do justice to the harm the phenomenon inflicts not only on targeted individuals, but also to the democratic and inclusive quality of the political domain at large in this region of the world. While a large corpus of research, both EU-related and country-specific, exists on general abuse and violence towards women politicians, its digital spillover has not been adequately addressed, despite the large evidence brought forward by many women participating in European political life both at the national and transnational level.

Nonetheless, individual studies and interviews involving both women politicians directly as well as their employees conducting their communication campaigns shed some light on the phenomenon and the way it has been framed, handled and occasionally brought up within their institutions. A study on the online campaigns of Theresa May and Jeremy Corbin in the UK, for instance, highlighted that the former had more than three times the share of comments of the latter

aimed at criticising her physical appearance, and even when it came to the two politicians' own posts on social media, responses to May's statements were significantly more negative than Corbyn's (Barboni, 2018). When looking at these replies, a gendered pattern emerged, whereby men accounted for 3.4 times as many replies to May than women (*ibidem*). As underlined by expert Mona Lena Krook, many of these comments fell under the guise of criticism aimed at May's policies, but even when they did address policy issues, they were nonetheless couched in gendered terms: "They're not saying, 'you're stupid', but 'you're a stupid woman'. It gets to the idea that as a woman you have no right to say something. It's that they feel affronted that it's a woman who is taking the space" (*ivi*, p. 30-31).

Attacks against women of colour and religious minorities are even more vicious, in Europe as well as in other areas of the world. A study by Amnesty International, carried out in the UK, found that Black and Asian women Members of Parliament (MPs) in the UK received 35% more abusive tweets than their white counterparts (Amnesty International, 2017). Of those abusive tweets, half were directed against Labour Party MP and Shadow Home Secretary Diane Abbott, whose experience has been "characteristically racist and sexist", including death threats and an English Defence League-affiliated Twitter account that launched the hashtag #burnDianeAbbot (*ibidem*). The intersectional approach necessary to understanding and addressing this type of abuse rests on a number of cases, such as Tasmina Ahmed-Sheikh, former MP for the Scottish National Party and the first Black Asian Minority Ethnic woman from Scotland to be elected to any Parliament, who explained the toll that online abuse takes on her; as well as British Labour Party MP Luciana Berger, who has been the subject of both misogynistic and anti-Semitic attacks, several of which have risen to the level of credible death threats (*ibidem*). To confirm the cyclical nature of these vicious attacks, Berger's decision to report the abuse and pursue the prosecution of the perpetrators led to even more anti-Semitic responses (*ibidem*).

The use of large masses of bots to attack women candidates and politicians, which dramatically came to the limelight during Hillary Clinton's 2016 US presidential elections campaign, is not foreign to Europe either: just days after setting up her Twitter account, former German Chancellor Angela Merkel was subjected to a wave of abuse from Russian trolls posting derogatory and sexualised comments and threats (*ibidem*). One particular picture, representing Merkel and former Ukrainian president Petro Poroshenko, became a target for comments comparing both leaders to Nazis, insulting Merkel's appearance, and making aggressive sexual threats (*ibidem*).

More recent ICTs have also introduced the possibility to modify different types of media content, including pictures and videos, constituting just another means for anonymous users to target and ridicule prominent women politicians. The president of Croatia, Kolinda Grabar-Kitarović, experienced this innovative type of abuse on her own skin, declaring that, around two days after the election, she found "movie clips from […] a porn movie […] claiming it [was her] 'in action'" (*ivi*, p. 33). To this day, "when [she] googles her name, [… she sees] all sorts of photos that [are] actually not [her], especially in bathing suits", something that "makes you feel like an object, rather than as an actor" (*ibidem*).

Even when not sexual in nature, gendered comments that pose double standards for female and male political leaders can also have a damaging impact, something British MP Rachel Reeves experienced in 2015, when she was pregnant, shortly before the parliamentary elections. While "lots of male MPs had children whilst being in the cabinet", "no one would have questioned their ability to do their job", something for which she, on the other hand, was largely criticised (*ibidem*). What was most surprising to Reeves was that "people, even if they thought it, would then be willing to say it", and "[it] was not just online", but also "on radio phone-ins and newspapers" (*ibidem*).

Responding to these vicious and pervasive attacks is not an easy task, for any action taken on the part of the target carries its own negatives. In most cases, it can hardly be prevented, too, as demonstrated by British Labour Party MP Jess Phillips's case, who decided to put filters on her social media account after reaching a peak of more than 600 rape and death threats in a single day: the filters still fail at shielding her from the constant and demeaning comments about her appearance, with people "find[ing] the most unattractive pictures" to "pull apart [one's] appearance", an uncommon occurrence with "craggy faced and tired" men politicians (*ivi*, p. 36). Women's appearance and the way it is used as a weapon against their credibility encompasses other aspects, including regional accents, for which they are deemed "common" and "stupid" or they overall attitude, including the way they stand or sit. To make it simple, every single aspect of a woman's looks and behaviour is constantly under scrutiny, something which "affects [their] self-esteem more than it changes [their] behaviour", making targets "think that way about [themselves]" through some type of "confirmation bias" (*ibidem*).

In the most serious cases, online harassment escalates to the level of physical threat, something which makes the targets legitimately fear for their own safety: the Speaker of the Italian Chamber of Deputies, Laura Boldrini, had to face this issue when she faced numerous rape and death threats from right-wing groups referring to her stances on immigration (*ibidem*). One of the threats following the death of a woman and the arrest of three African immigrants for her murder was her face photoshopped onto a severed head alongside the comment "This is the fate she must endure in order to appreciate the customs of her friends". In the aftermath of this post, which received multiple responses on the same wave, Boldrini stated that the event "really affect[ed her] a lot", with some users going as far as saying that "[they knew] where [her] apartment [was]" as well as "[her] address" (*ivi*, p. 37). These latter instances also made Boldrini fear for her 20-year-old daughter studying in the UK, and she feels guilty and anxious that she may "have a problem because of [her mother]" (*ibidem*).

Many more European interviewees commented on the damaging impact their experience and the measures they have to take to protect themselves have on the European democratic health at large. Anne Jenkin, a member of the UK House of Lords, underlined that "the most powerful thing a politician can do is be themselves, and it is their stories and their backstories that makes it interesting", but the response women politicians receive "inhibits [them] from telling that story, because they do not want to personalise themselves, and of course they are worried about their families" (*ivi*, p. 34). This legitimate de-personification is a shame, however, "because we have a huge variety of backgrounds of people in politics", an element which is particularly fundamental for democratic systems, based on inclusion and dialogue among different types of people, "and yet they do not want to talk about where they come from for that reason" (*ibidem*). More specifically, women face a double-edged sword when it comes to share their personal side, because "one of the biggest challenges" – according to Jenna Lowenstein – "is their need to credential themselves", for "people do not take their experience […] and […] occupation seriously" (*ibidem*). However, by putting the accent on their achievement, women end up making "people think [they] are all resume, no soul", a risk which causes them to be stuck "walking the line between that necessary work of credentials and [at the same time] not coming across as only accomplishments" (*ibidem*). This difficult limbo has come up in numerous studies showing how "women communicating in a mitigated and less competent style were better able to influence men than women using a more competent style" as "men, but not women, reported that highly competent women were more

threatening and less likeable than less competent women, and these negative perceptions reduced [their] influence" (Linda, 2001). This double standard further complicates women's political communication, for qualities that seem, at surface level, incredibly desirable for politicians, often end up backfiring when attributed to women, an example being the ability of going unscripted: in the case of a man, he can easily come off as rogue and funny, but when it is a woman, she is likely to be deemed unprepared and rude (*ibidem*).

Despite the many trigger events and individual experiences of online abuse against women politicians in Europe, only the most recent reports address the specific of violence towards them in cyberspace. One of the reports presenting disaggregate data at the national level within the European Union was published in February 2024, stating that, for one, 90% of all female deputies in the German Bundestag have been the target of hate speech, 57% of them on sex-related grounds; whereas one study on the Irish Parliament reports that 96% of those interviewed had received threatening social media or email messages or 'hate mail'; while three quarters said they had been threatened online with physical violence (European Parliament, 2024).

3.4 European measures against (online) violence towards women in politics

European and other international measures against online violence towards women as a phenomenon make up a large corpus of conventions and recommendations. However, the specific issue of online violence against women in politics as a differentiated type of violence, which does not only harm the subject but its position and its links with the democratic process at large, have rarely been addressed in the European landscape.

Part of the issue can be resolved by extending existing measures against violence towards women in politics. The European Union, for one, has had, since 2002, a Directive on Equal Treatment that "enforces equal working conditions between women and men and defines sexual harassment as sex discrimination and a violation of dignity" (Berthet & Kantola, 2021, p. 31: Zippel 2006, 2008). Because this Directive is legally binding, Member States had to comply. Considering that the European Union had no competence over "violence against women" as a policy field, many scholars saw this Directive as one of European feminists' major achievements (Kantola 2010; Berthet & Kantola, 2021, p. 146: Zippel, 2008). In fact, the Directive's adoption was enables by

advocates for an EU-wide intervention against sexual harassment framing it strategically as a workplace issue (Berthet & Kantola, 2021: Zippel, 2008). Using the 1976 Equal Treatment Directive, which provided equal treatment for men and women in the workplace, including working conditions, the measures were extended to the protection from and prohibition of sexual harassment (Berthet & Kantola, 2021: Zippel, 2008).

Additionally, the political work of the European Parliament is guided by its Rules of Procedure, which have been frequently modified to ensure their effective working. In 2006 and 2017, reforms to the Rules of Procedure introduced new penalties for MEPs and staff engaging in inappropriate behaviour, including Rule 11 which bans defamatory, racist, and xenophobic language or behaviour. However, penalties were controversial and rarely applied (Berthet & Kantola, 2021: Brack, 2017). Moreover, these reforms did not specifically address sexual harassment, and the issue, as well as the procedures related to it, remained undefined (*ibidem*).

In terms of regulation of online violence against women, the Committee on Standards in Public Life assessed the steps that the UK government should take to better tackle the issue, many of which could lay the basis for similar measures in other countries (Barboni, 2018). One of their key recommendations is to "shift the liability of illegal content online towards the social media companies", noting that "currently, social media companies do not have liability for the content on their sites, even where that content is illegal" (*ivi*, p. 47). This is largely due to the EU E-Commerce Directive (2000), which treats the social media companies as 'hosts' of online content. It is clear, however, that this legislation is out of date. Facebook, Twitter, and Google play a role in shaping what users see. […] The time has come for the companies to take more responsibility for illegal material that appears on their platforms" (*ibidem*).

National governments such as Germany also attempt to "sentence social media companies that do not act swiftly to remove abusive content, imposing a hefty fine (€500,000) for each item of hate speech that remains on a social media platform for over 24 hours" (*ivi*, p. 48). Other governments are also weighing new measures to punish perpetrators, with the UK's Committee on Standards in Public Life recommending "the introduction of a new offence in electoral law of intimidating Parliamentary candidates and party campaigners" (*ibidem*). Additional measures, specifically tailored for the issue applied to politicians, "have focused on protecting candidates' personal information, in order to help prevent threats from turning into physical intimidation or violence"

(*ivi*, p. 48). For example, the Committee has urged the government to "bring forward legislation to remove the requirement for candidates standing as local councillors to have their home addresses published on the ballot paper" (*ibidem*).

Moreover, in the UK, the Crown Prosecution Service (CPS) published, in 2016, new rules aimed at stepping up prosecution of online harassment. Cyber offences that could result in criminal charges according to the new legislation include creating hashtags and trends that encourage online harassment, spreading or encouraging the spread of offensive images, doxing, and posting "disturbing or sinister" photoshopped images (*ibidem*).

The Metropolitan Police's efforts have also spiked in the aftermath of the murder of Labour MP Jo Cox, establishing a Parliamentary Liaison and Investigation Team (PLAIT) charged with responding to threats against MPs. The main issue is the effort and resources needed to deal with the "copious" numbers of threats: in its first year alone, the PLAIT team dealt with 102 complaints (*ibidem*).

Additionally, measures have been adopted in some countries in the European Union to handle verbal attacks against women politicians within political debate, setting out some useful groundwork for extending the principles and recommendations to the cyberspace: in Finland, Section 31 of the Constitution (731/1999) (p. 7) states that "representatives must conduct themselves with dignity and decorum, and not behave offensively to another person" (Constitution of Finland, 2000). In France, Article 70 of the Rules of Procedure of the National Assembly penalises certain types of behaviour in public sittings, such as "attacks, insults, provocations or threats" against colleagues (Rules of Procedure of the National Assembly, 1999). In the German Bundestag, members must "respect the dignity of the house during debates" (Rules of Procedure of the German Bundestag and Rules of Procedure of the Mediation Committee, 2022). Similar measures exist in Luxembourg, Portugal, Romania and others. Additional regulations concern attacks, threats, and harassment outside of debate, one example being represented by the Finnish parliament, where representatives "receive guidance on cyber-harassment, including on information security and on how to deal with any harassment or intimidation, as well as receiving psychological support from the medical service" (European Parliament, 2024, p. 10). The strengthening of support and intervention of this kind within political institutions has been recently improved in European countries, including Germany, the Netherlands, Poland, and Sweden. And

while not all of them directly address the pervasiveness of targeted online abuse, the furthering of institutional protection of representatives belonging to minorities and vulnerable groups holds the promise of more extensive norms encompassing the political dimension of cyberspace.

3.5 Gaps in research and regulation

Despite the sheer and concerning volume of online harassment towards women in public life within the European borders and specific member states, no existing research on the issue qualifies as satisfactory and in terms of aggregate data or ways forward which may be applied on a large enough scale. Gaps in regulation emerge both at the national and transnational level, partially – but not exhaustively – addressing the issue of gender-based harassment towards women in politics, but rarely focusing to a sufficient extent on the specific and unique dimension of online abuse.

As this work proves, it is absolutely not sufficient to research online violence towards women in politics and in prominent public positions more in general by treating the phenomenon as isolated, for its roots reside in antecedent gender-roles, power dynamics, and intersectional discrimination of which online abuse towards women in politics is merely the more evident and transversal epitome. Addressing gender inequalities not only in composition, but also in substantial access and representation of women and especially minority women within political institutions is strictly necessary to lay the basis for a wider discussion, encompassing civic engagement both in grassroot political activism and the electoral domain as well. Until we do not address the larger issue of women's involvement at all levels of public and political life, we will not be able to ensure their protection, democratic participation, and adequate representation in institutions. And unless this more pragmatic basis is reinforced, we will hardly manage to protect them both as individuals as well as as public figures from any kind of harassment, especially that occurring in the borderless, anonymous, and often non-prosecutable environment represented by the digital domain.

Existing measures regulating online violence towards women and workplace harassment towards women in politics thus need to converge, at the transnational level, in order to cover this widely forgotten gap, in a multilateral effort to promote the potential of cyberspace as a milieu of democratic inclusion, representation and debate.

4. CONCLUSION

Ultimately, the hopeful aim of this work is to observe the phenomenon of violence against women in the public sphere from an encompassing, multilayered perspective, allowing to detect the different levels at which action must be taken for a significant change to occur.

In the first part, pulling the strings of three different academic domains has allowed me to build an original top-down paradigm, framing the issue in normative, political, and social terms. From the top, digital constitutionalism provides excellent groundwork on the fundamental need for a normative approach to cyberspace as a domain of potential enhancement of democratic quality, human rights, and ultimately, to the aim of this work, gender inclusion.

Using this framework, I have turned to the analysis of online abuse, analysing current regulations and controversies around it, and especially the impact it has on the free political expression of women in the public sphere. Such a phenomenon, still viciously present, calls into question the normative framework at the basis of the structure and functioning of cyberspace, especially in its declination as a public platform and forum, extending the offline public sphere in a unique, borderless, and interactive manner. For cyberspace to provide a positive, additional dimension to the public sphere and political debate, rather than representing a dangerous, lawless forum, a digital constitutionalist approach recommends the enhancement and protection of human rights and freedom of expression on the part of disadvantaged groups, including women. The political dimension of public debate in cyberspace is all the more relevant in terms of the opportunities it poses to disenfranchised public figures, including women, who have the chance to create their own direct channels for political communication on their platforms, proving the point that cyberspace holds the potential for democratic innovation thanks to its unique, unprecedented characteristics, which have no equal in the offline, traditional system. Reflections on existing democratic innovations, then, provide fundamental insights as to what needs to be done to enhance this innovative character of cyberspace, while carefully addressing the risks and dangers that such an ambiguous tool also carries. Finally, the focus on technofeminism provides the unifying element of the gender-based perspective, which frames the normative approach in terms of constitutional norms applied to gender equality, equal opportunity, intersectionality, and the overcoming of the masculine paradigm of political power; and encases reflections on digital democratic innovations

in terms of enhancing the cyberspatial potential to provide an original, inclusive, and gender-sensitive political space for women in the public sphere to create their own narrative, express their perspective and opinions, and insert themselves in an otherwise traditionally masculine domain. Addressing the issue of online violence towards women in the public sphere is, in a sense, the first step, in that it encompasses the necessity for a normative framework protecting women's rights online; the importance of these rights as driving elements of democratic quality; and the fundamental gender-related characterisation of the phenomenon, which links it to pre-existent cultural paradigms that need to be overcome.

As an approach that counts many European authors, digital constitutionalism can be applied to the European landscape when analysing existing transnational norms stating the normative intention to shape cyberspace as an inclusive and pluralistic space, protecting and enhancing human and citizens' rights both online and offline. Data protection and other controversial issues have been addressed by multiple European regulations, such as the Digital Data Protection Regulation (2018), and the issue of online violence against women and other minorities has been addressed in recent years, drawing on international research and norms. Framing the European debate and evolution from a gender perspective is fundamental in resorting to the root causes of persisting electoral and institutional violence towards women, rooted in the culturally masculine, exclusive, and elitist paradigm that still characterises the political domain at both the national and continental level. The pervasiveness of online violence against women representatives in Europe is a transposition of offline abuse, exclusion, and silencing of women within a public space that causes a spillover effect. Linking the phenomenon to a matter of democratic quality, of which inclusion and pluralism are main elements, highlights the urgency of the issue and the potential for democratic innovation represented by its addressal, especially online.

penso sarebbe buono che tu indicassi un paragrafo quali sono i liiti di questo lavoro (dovuti al tempo, risorse etc) e soprattutto indicassi alcune piste di ricerca esplicite. Non genericamente, ma dicendo: sarebbe importante sviluppare un filone che guardi a XXX con questa e questa metodologia XXX al fine di mettere in evidenza XXX; e poi anche analizzare XXX insomma: mostrare che il lavoro di 'ricognizione'che hai fatto ti ha fornito elementi per possibili futuri approfondimenti

Despite my attempt to bring together a wide range of literature on the various levels of discrimination against women in the public sphere, and especially within cyberspace, a lot of questions are left unanswered. The time and research constraints of my work, adding to the lack of extensive research on the specificity of online violence towards women in politics, leave the door open for further reflections and empirical studies: what I could do, within the limits of my knowledge and tools, was to craft an original theoretical approach to the issue, addressing its normative requirements and democratic impact.

First and foremost, the elephant in the room remains: women are systematically silenced, one way or the other, when they enter the public sphere. An alarming number of female MPs have stepped down, for instance, in the UK, openly mentioning the institutional and public abuse they experienced as one of the reasons. In line with above-mentioned studies, most of them are young and have been in parliament for less than a decade (Parraudin & Murphy, 2019). Similar events have been observed within European institutions, as mentioned above. While most reports do not fully address the role of online violence, many representatives directly refer to it, explaining how it led them to either limit their platforms, filter interactions, close them off completely, or simply attempt not to touch on controversial issues that could backfire. Regardless of how women are silenced and what their reaction is, political violence against them, especially online, profoundly limits their ability to fully participate in the democratic process, speak their mind, and stand their ground; for even in those cases were they react fiercely and keep doing what they do, the psychological and emotional impact is non-negligible. Therefore, there remains a research vacuum in determining the disaggregated proportions of online violence experienced by women politicians, and the silencing effect it has on them, damaging their career, the electorate they represent, and democratic quality at large. Quantitative research providing more precise disaggregate data on the issue would allow for a more comprehensive study of the phenomenon and its unique consequences for democratic quality.

The more practical question, then, is how it is possible to disaggregate data concerning the online harassment experienced by women in politics be tracked, documented and addressed. This is undoubtedly the most necessary step to take in order to address the pervasiveness and viciousness of the problem, which has already been widely acknowledged. Quantitative studies appealing directly to the targeted subjects within different institutions would paint a more comprehensive

picture, potentially enabling comparative research among different institutions and political realities.

The follow-up question concerns the levels at which the issue can be addressed: much data exists on the multimodal harassment and abuse of women in politics, especially within major political institutions, both at the national and transnational level, but the online spillover, whose unique characteristics make it non-regulable through traditional norms, is scarcely researched. While the social and cultural root causes are clear, and do not differ that much from the sexist abuse women experience more broadly in society, the unique status and role of women in the public sphere adds a layer of complexity to the political and social consequences of the violence they experience. The fact that, in face of this pervasive issue, public authorities and existing legislation currently fail to present an exhaustive framework for its addressal, undoubtedly constitutes an extremely problematic regulative context. The insights provided by literature on digital constitutionalism constitute an important stepping stone in addressing the spillover abuse in cyberspace, but the latter will hardly be safer and inclusive of women's democratic participation unless the offline dangers are socially and politically acknowledged.

The necessity for a feminist framework is all the clearer given that existing regulation on hate speech hardly adopts a gender perspective, failing to account for the intersectional nature of the issue, especially when it comes to a complex and unique type of abuse that specifically targets women in charge in the public sphere. Drawing on insights provided by cyberfeminism and technofeminism, innovative reflections can be carried out on the gendered nature of online abuse, as well as – on a lighter note – on the potential that cyberspace still holds for inclusion, if its limitations and danger are adequately addressed. This is all the more important if we consider the political forum represented by cyberspace, which constitutes nowadays a fundamental component of civil society and thus of public discourse.

All in all, the wider question would then be whether this complex work of acknowledgement and addressal of gender abuse in the public sphere and within its prolongment constituted by cyberspace can eventually fulfil the promises of the latter as a democratic innovation, enlarging the possibilities for inclusive democratic participation to public debate.

The many controversies discussed in this thesis highlight the dichotomous nature of the public sphere as both a threatening reality for authoritative women, as well as a potential world of

possibilities for emancipation. The unique nature of cyberspace further deepens this contradiction, urging all the competent actors to come together to contain the damage and enhance its enormous potential. While I would never claim to know the answers to any of these questions, the hope is that bridging different approaches and aspects can provide meaningful insights as to what needs to be done at different levels to realise the techno-feminist hope that an enhancement of the public sphere aware of the layers of abuse and discrimination eventually turns the tide around, providing an unprecedentedly inclusive space for democratic growth.

4. BIBLIOGRAPHY

Ahrens P. (2012), *Wenn sich Nachteile als Vorteile erweisen: Der Ausschuss für die Rechte der Frau und die Gleichstellung der Geschlechter im Europäischen Parlament*. Femina Politica

Amaral I., Simões R. B. (2021), *Violence, Misogyny, and Racism: Young Adults' Perception of Online Hate Speech*. University of Coimbra

Amnesty International (2018), *Toxic Twitter - a Toxic Place for Women*

Anderson C. W. (2011), *Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience Transparency*. International Journal of Communication

Armed Conflict Location & Event Data Project (2021), *Violence targeting women in politics: trends in targets, types, and perpetrators of political violence*

Asenbaum H. (2020), *Making a Difference: Toward a Feminist Democratic Theory in the Digital Age*. University of Westminster

Banks K. (2001) *Leave the Internet Alone*. APWIN

Barboni E. (2018), *(Anti)social Media: The Benefits and Pitfalls of Digital for Female Politicians*. Atalanta

Bardall G. (2013), *Gender-Specific Election Violence: The Role of Information and Communication Technologies*. Stability: International Journal of Security & Development

Bartlett J. (2014), *Social media is transforming how to study society*. Demos

Beauvais E., Beachtiger A. (2016), *Taking the Goals of Deliberation Seriously: A Differentiated View on Equality and Equity in Deliberative Designs and Processes*. Journal of Public Deliberation, vol. 12

Becker G. (1991), *A Treatise on the Family*. Harvard University Press

Benkler Y. (2006), *The wealth of networks: How social production transforms markets and freedom*. Yale University Press

Bennet W. L., Livingston S. (2020), *The disinformation age: Politics, technology, and disruptive communication in the United States.* Cambridge University Press

Benton et al. (2022), *Hate Speech Spikes on Twitter After Elon Musk Acquires the Platform.* School of Communication and Media Scholarship and Creative Works

Berg S., Hofmann J. (2021), *Digital Democracy.* Internet policy review, vol. 10, issue 4

Berman P. S. (2000), *Cyberspace and the State Action Debate: The Cultural Value of Applying Constitutional Norms to 'Private' Regulation.* GW Law Faculty Publications & Other Works

Berthet V., Kantola J. (2021), *Gender, Violence, and Political Institutions: Struggles over Sexual Harassment in the European Parliament.* Social Politics, vol. 28, no. 1

Bertolini G. (2022), *A feminist digital constitutionalism?* University of Padua

Binns A. (2017), *Fair Game? Journalists' Experience of Online Abuse,* Journal of Applied Journalism and Media Studies

Boeckmann R. J., Turpin-Petrosino C. (2002), *Understanding the Harm of Hate Crime.* Journal of Social Issues, vol. 58, no. 2

Brack N. (2017), *Opposing Europe in the European Parliament: Rebels and radicals in the chamber.* Basingstoke: Palgrave Macmillan

Bray F. (2007), *Gender and Technology.* Annual Reviews

Breslin B. (2009), *From Words to Worlds: Exploring Constitutional Functionality.* The Johns Hopkins University Press

Brophy J. (2010), *Developing a Corporeal Cyberfeminism: Beyond Cyberutopia,* New Media and Society

Byrne J. P. (1991), *Racial Insults and Free Speech Within the University.* Georgetown Law Faculty Publications and Other Works

Carastathis A. (2016*), Intersectionality: Origins, Contestations, Horizons.* University of Nebraska Press

Carlson C. R. (2017), *Misogynistic hate speech and its chilling effect on women's free expression during the 2016 US presidential campaign.* Journal of Hate Studies, vol. 14, issue 1

Casarosa F. (2020), *Freedom of Expression and Hate Speech in the Framework of the 'E-Learning National Active Charter' Project.* European University Institute

Celeste E. (2019), *Digital constitutionalism: a new systematic theorisation.* International Review of Law Computers & Technology

Celis et al. (2014) *Constituting Women's Interests through Representative Claims.* Politics & Gender

Chiva C. (2014), *Gender, European Integration and Candidate Recruitment: The European Parliament: Elections in the New EU Member States.* Parliamentary Affairs

Citron D. K., Norton H. (2011), *Intermediaries and Hate Speech. Fostering Digital Citizenship for our Information Age.* University of Colorado School of Law

Coleman S. (2017), *Can the internet strengthen democracy?* Polity Press

Cockburn C. (1983), *Brothers: Male Dominance and Technological Change.* Pluto Press

Cohen-Almagor R. (2011), *Fighting Hate and Bigotry on the Internet.* Policy and Internet

Cole D. (1996), *Racist speech should be protected by the constitution.* Greenhaven

Cowan R. S. (1976), *The 'Industrial Revolution' in the Home: Household Technology and Social Change in the Twentieth Century.* Technology and Culture, vol. 17

Curato et al. (2017), *Twelve Key Findings in Deliberative Democracy Research.* The American Academy of Arts & Science

Dahlgren P. (2005), *The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation.* Political Communication

Dalton et al. (2003), *Democracy Transformed? Expanding Political Opportunities in Advanced Industrial Democracies*. Oxford University Press

Daniels J. (2009), *Rethinking Cyberfeminism(s): Race, Gender, and Embodiment*. The Feminist Press at the City University of New York

De Beauvoir S. (1949), *The Second Sex*. Editions Gallimard

De Blasio E., Viviani L. (2020), *Platform Party between Digital Activism and Hyper-Leadership: The Reshaping of the Public Sphere*. Media and Communication

Delgado R., Stefancic J. (1994), *Hateful Speech, Loving Communities: Why Our Notion of "A Just Balance" Changes So Slowly*. California Law Review

De Silva A. (2021), *Hate Speech Against Women: Addressing a Democratic Crisis*. Melbourne School of Government

De Zúñiga et al. (2010), *Digital Democracy: Reimagining Pathways to Political Participation*. Journal of Information Technology and Politics

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

Dworkin R. (1986), *Law's Empire*. Harvard University Press

Erjavec K., Kovačič M. P. (2012), *Hate Speech: A Systematized Review*. Sage Publications

EurActiv (2014), *Lack of women no longer a problem for new Commission*

European Federation of Journalists (2018). *About #mediaagainsthate*

European Commission (2016), *Code of conduct on countering illegal hate speech online*: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

European Council (2006), *Council Decision*, 2006/515/EC: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32006D0515

European Council (2008), *Council Framework Decision*, 2008/913/JHA: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008F0913

European Parliament (2009), *Charter of Fundamental Rights*, 2012/C 326/02: https://www.europarl.europa.eu/charter/default_en.htm

European Parliament (2023), *Women in politics in the EU State of play*: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739383/EPRS_BRI(2023)739383_EN.pdf

European Union (2016), *EU Code of Conduct*: https://commission.europa.eu/about-european-commission/service-standards-and-principles/ethics-and-good-administration/commissioners-and-ethics/code-conduct-members-european-commission_en#:~:text=The%20Code%20of%20Conduct%20provides,wide%20politics%20and%20national%20politics.

European Union (2016), *General Data Protection Regulation*, EU Regulation no. 2016/679: https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/#:~:text=The%20GDPR%20establishes%20the%20general,data%20processing%20operations%20they%20perform.

European Union (1953), *European Convention on Human Rights*: https://www.echr.coe.int/european-convention-on-human-rights

Facebook (2019), *Oversight Board*

Fernandez M. (2003), *Cyberfeminism, Racism, Embodiment*. Autonomedia

Finkel et al. (2020*), Political sectarianism in America.* Science 370

Fiss, O. (1996), *The irony of free speech.* Harvard University Press

Fitzgerald B. (1999), *Software as Discourse? A Constitutionalism for Information Society*. Alternative Law Journal

Fitzgerald B. (2000), *Software as Discourse? The Challenge for Information Law*. European Intellectual Property Review

Fortin-Rittberger J. (2014) *Do electoral rules matter? Explaining national differences in women's representation in the European Parliament*. European Union Politics

Frenkel S., Conger K. (2022), *Hate Speech's Rise on Twitter Is Unprecedented, Researchers Find.* New York Times

Franklin M. I. (2021), *Digital suffragists: women, the web, and the future of democracy. By Marie Tessier.* Goldsmiths University of London

Fraser N. (1990), *Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy.* Duke University Press

Froomkin M. (2004), *Technologies for Democracy.* Democracy Online: The Prospects for Political Renewal Through the Internet

Garcia, L., & McDevitt, J. (1999), *The psychological and behavioral effects of bias and non bias motivated assault.* U.S. Department of Justice, National Institute of Justice

Gerstenfeld, P. B. (1992), *Smile when you call me that! The problems with punishing hate motivated behavior.* Behavioral Sciences and the Law

Gelber K., Stone A. (2007), *Hate Speech and Freedom of Speech in Australia.* Federation Press

Gill R. (2007), *Gender and the Media.* Polity Press

Global Media Monitoring Project (2015). *U.S. national report*

Grigg, W. N. (1992), *The hate-crime label is used to suppress politically incorrect opinions.* Greenhaven Press

Hamm M. S. (1994), *Conceptualizing hate crime in a global context.* Anderson Publishing and Academy of Criminal Justice Sciences

Haraway D. (1985), *The Cyborg Manifesto.* Socialist Review

Hardaker C., McGlashan M. (2015), *''Real men don't hate women'': Twitter rape threats and group identity.* Journal

Harding S. (1986), *The Science Question in Feminism.* Cornell University Press

Herek et al. (2002), *Victim experience in hate crime based on sexual orientation.* Journal of Social Issues

Herring S. C. (2003), *Gender and power in online communication.* Blackwell

Hickey et al. (2023), *Auditing Elon Musk's Impact on Hate Speech and Bots.* Seventeenth International AAAI Conference on Web and Social Media

Holton A. E., Molyneux L. (2015), Identity lost? The personal impact of brand journalism. *Journalism*

Howard P. N. (2020), *Lie machines: How to save Democracy from troll armies, deceitful robots, junk news operations, and political operatives.* Yale University Press

Hubig C. (2006), *Die Kunst des Möglichen: Grundlinien einer dialektischen Philosophie der Technik.* Transcript

Hughes T. P. (1986), *Technological Momentum.* MIT Press

International Women's Media Foundation (2014), *Violence and Harassment against Women in the News Media: A Global Picture*

Inter Parliamentary Union (2018), *Sexism, harassment and violence against women in parliaments in Europe*

Jacobs, J. B., & Potter, K. (1998), *Hate crimes: Criminal law and identity politics.* Oxford University Press

Jaffee S., Hyde J. S. (2000), *Gender differences in moral orientation: A meta-analysis.* Psychological Bulletin

Jon, G. B. (2005), *Speak No Evil: The Triumph of Hate Speech Regulation.* University of Chicago Press

Kantola J. (2010), *Gender and the European Union.* Basingstoke: Palgrave

Kantola J., Nousiainen K. (2012), *The European Union: Initiator of a New European Anti-Discrimination Regime?* Krizsan

Kendall R. (1996), *'MUDder? I Hardly Knew 'Er!': Adventures of a Feminist MUDder.* Seal Press

Klepper D. (2023), *Musk threatens to sue researchers documenting the rise in hateful tweets.* PBSO News Hour

Kravas V. (2010), *Governance of Virtual Worlds and the Quest for a Digital Constitution*. Edward Elgar Publishing

La Paz et al. (2020), *Hate Speech: a Systematized Review*. Sage Publications

Lawrence F. M. (1999), *Punishing hate: Bias crimes under American law*. Harvard University Press

Landwehr C. (2015), *Democratic Meta-Deliberation: Towards Reflective Institutional Design*. Political Studies Association, vol. 63(S1)

Lee J. (2015), *The Double-Edged Sword: The effect of journalists' social media activities on audience perceptions of journalists and their news products*. Journal of Computer-mediated Communication

Lohan M. (2000), *Constructive Tension in Feminist Technology Studies*. Sage Publications

Lühiste M., Banducci S. (2016), *Invisible Women? Comparing Candidates' News Coverage in Europe*. Cambridge University Press

MacKinnon C. A. (1993), *Only words*. Harvard University Press

MacKinnon C. A. (2017), *Butterfly politics*. The Belkap Press of Harvard University Press

Macrae H. (2010) *The EU as a Gender Equality Policy: Myths and Realities*. Journal of Common Market Studies

Mann M. (2018), *The limits of (digital) constitutionalism: Exploring the privacy-security (im)balance in Australia*. The International Communication Gazette

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

Massaro T. M. (1991), *Equality and Freedom of Expression: The Hate Speech Dilemma*. William & Mary Law Review, vol. 32

Matsuda et al. (1993), *Words that wound: Critical race theory, assaultive speech, and the First Amendment.* Westview

Meyer H., Carey M. C. (2014), *In Moderation: Examining how journalists' attitudes toward online comments affect the creation of community.* Journalism Practice

Mushaben J. (1998), *The Politics of Critical Acts: Women, Leadership, and Democratic Deficits in the European Union.* The European Studies Journal

Musk E. (2022), X: https://twitter.com/elonmusk/status/1499976967105433600?s=20

National Democratic Institute, Ipsos (2014), *Women, Technology, and Democracy Survey.* National Democratic Institute

Neumann, E.N. (1974), *The spiral of silence: A theory of public opinion.* Journal of Communication

Nguyen et al. (2008), *Moral issues and gender differences in ethical judgment using Reidenbach and Robin's (1990) multidimensional ethics scale: Implications in teaching of business ethics.* Journal of Business Ethics

No Hate Speech Movement (2018), *About the Campaign*

O'Brien J. (1997), *Changing the Subject.* New York: Women and Performance Project

O'Brien M., Ortutay B. (2022), Musk's Twitter disbands its Trust and Safety advisory group. *Associated Press*

Organisation for Security and Cooperation in Europe (2021), *Violence Against Women Journalists and Politicians: a Growing Crisis*

O'Sullivan D. (2022), *Twitter is no longer enforcing its Covid misinformation policy.* CNN Business

Oudshoorn N., Pinch T. (2003), *Users Matter: The Co-Construction of Users and Technology.* MIT Press

Padovani C., Santaniello M. (2018), *Digital constitutionalism: Fundamental rights and power limitation in the Internet eco-system.* The International Communication Gazette

Parraudin F., Murphy S. (2019), *Alarm Over Number of Female MPs Stepping Down After Abuse.* The Guardian

Paul J. (1991), *The Hidden Structure of Takings Law*. Southern California Law Review, vol. 64

Phillips A. (1995), *The Politics of Presence*. Oxford University Press

Phillips W. (2012), *What an Academic Who Wrote Her Dissertation on Trolls Thinks of Violentacrez*. The Atlantic

Pitts V. (2004), *Illness and Internet Empowerment: Writing and Reading Breast Cancer in Cyberspace*. Health: An Interdisciplinary Journal for the Social Study of Health, Illness, and Medicine

Plant S. (1997), *Zeros + Ones*. 4th Estate Limited

Posetti J., Shabbir N. (2022), *The Chilling: A global study of violence against women journalists*. International Center for Journalists

Post D. G. (2000), *Of Black Holes and Decentralized Law Making in Cyberspace*. Vanderbilt Journal of. Entertainment and Technology Law 70

Praiser E. (2011), *The filter bubble: What the Internet is hiding from you*. Viking

Prentice D. A., Carranza E. (2002), *What men and women should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes*. Psychology of Women Quarterly

Radin M. J., Wagner P. (1999), *The Myth of Private Ordering: Rediscovering Legal Realism in Cyberspace*. Chicago-Kent Law Review

Raikka J. (2003), *The argument from truth as a basis for legislation concerning freedom of speech*. Trames

Renman V., Conroy C. (2014) *Advances in EU Gender Equality: Missing the mark?* EPIN Paper Nr. 41

Şahi et al. (2018), *Automatic Detection of Hate Speech Towards Women on Twitter*. 3rd International Conference on Computer Science and Engineering

Sassen S. (2002), *Towards a Sociology of Information Technology*. Sage Publications

Solovev K., Pröllochs N. (2022), *Hate Speech in the Political Discourse on Social Media: Disparities Across Parties, Gender, and Ethnicity*. Cornell University

Southworth et al. (2007), *Intimate partner violence, technology, and stalking*. Violence Against Women

Stanley A. (1995), *Mothers and Daughters of Invention*. Rutgers University Press

Sustein C. R. (2017), *#Republic: Divided democracy in the age of social media*. Princeton University Press

Suzor N. (2010), *Digital Constitutionalism and the Role of the Rule of Law in the Governance of Virtual Communities*. Queensland University of Technology

Tessier M. (2021), *Digital suffragists: women, the web, and the future of democracy*. MIT Press

Teubner G. (2012), *Constitutional Fragments: Societal Constitutionalism and Globalization*. Oxford University Press

Tsesis A. (2002), *Destructive messages: how hate speech paves the way for harmful social movements*. New York University Press

Turkle S. (1984), *The Second Self: Computers and the Human Spirit*. Granada

Turpin-Petrosino C. (1999), *Connecting the past to the future: Hate crime in America*. Journal of Contemporary Criminal Justice

UK Parliament (2021), *Draft Online Safety Bill*

Uyheng J., Carley K. M. (2020), *Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines*. Journal of computational social science

Van Blarcum C. D. (2005), *Internet hate speech: The European framework and the emerging American Haven*. Washington & Lee Law Review

Verge T. (2020), *A chamber of one's own: institutional claim-making on gender equality and the symbolic role of parliaments*. Politics & Gender

Vidgen et al. (2021), *Understanding Online Hate*. Alan Turing Institute

Wajcman J. (2010), *Feminist theories of technology*. Oxford University Press

Wilding F. (1998), *Notes on the Political Conditions of Cyberfeminism*. Art Journal, vol. 57, no. 2

Wilhelm C., Joeckel S. (2018), *Gendered Morality and Backlash Effects in Online Discussions: An Experimental Study on How Users Respond to Hate Speech Comments Against Women and Sexual Minorities*. Sex Roles: A Journal of Research

Winkler et al. (2017), *Backlash effects in online discussions: Effects of gender and counter-stereotypical communication on persuasiveness and likeability.* International Communication Association

Wojatzki et al. (2018), *Do Women Perceive Hate Differently: Examining the Relationship Between Hate Speech, Gender, and Agreement Judgments*. University of Duisburg-Essen

Wojciechowska M. (2018), *Towards Intersectional Democratic Innovations*. Sage Publications

Women's Media Center (2017), *The status of women in the U.S*

Yilma K. (2017), *Digital privacy and virtues of multilateral digital constitutionalism – preliminary thoughts*. International Journal of Law and Information Technology

Young I. M. (1990), *Justice and the Politics of Difference.* Princeton University Press

Zannettou et al. (2018), *What is Gab: A bastion of free speech or an alt-right echo chamber*. WWW

Zippel S. K. (2006), *The politics of sexual harassment.* Cambridge University Press

Zippel S. K. (2008), *Violence at work? Framing sexual harassment in the European Union*. Gender politics in the expanding European Union: Mobilization, inclusion, exclusion