

Addressing the hate speech issue in the Nordic countries

A challenge for media welfare states or a chance for their revival?

REETA PÖYHTÄRI

FACULTY OF INFORMATIONTECHNOLOGY AND COMMUNICATION SCIENCES,
TAMPERE UNIVERSITY, FINLAND

ABSTRACT

Hate speech on digital platforms manifests itself in ways that need to be addressed both in practice and on a policy level, taking Nordic (media) policies into novel areas. For this chapter, I analysed and compared policy documents discussing hate speech from Denmark, Finland, Norway, and Sweden, with a focus on how hate speech is defined, the perceived harmful functions of hate speech, and the suggested remedies. The analysis shows that hate speech is perceived to endanger societies' democratic functions, public debate, and freedom of speech. Hate speech targets specific minorities and especially those in public positions or participating in public debate. The suggested remedies correspond with media welfare state ideals: increased collaboration between all relevant parties, state support for both citizens and media, and adjustments of existing laws. The role of online platforms is crucial, but concrete measures to hold them responsible still wait for the implementation of the European Union's Digital Services Act legislation.

KEYWORDS: hate speech, policy, rights and values, Nordic countries, digital media

Pöyhtäri, R. (2024). Addressing the hate speech issue in the Nordic countries: A challenge for media welfare states or a chance for their revival? In P. Jakobsson, J. Lindell, & F. Stiernstedt (Eds.), *The future of the Nordic media model: A digital media welfare state?* (pp. 71–94). Nordicom, University of Gothenburg. <https://doi.org/10.48335/9789188855893-3>

Introduction

Hate speech, especially in the online sphere, has become a great concern within Europe in the past decade. Hate speech as form of communication attacks various minorities and is used, for example, to spread ideas of racism, religious hatred, misogyny, and homophobia, but also to attack those in public positions, aiming to delegitimise their power through silencing (see, e.g., Council of Europe, 2023; Van Sant et al., 2021). Furthermore, hate speech puts at risk the communication and democratic values held especially high in the Nordic context: freedom of speech, access to information, open society, and equality (see, e.g., Carlsson & Pöyhtäri, 2017). These Nordic values are also connected to the principles of the media welfare state, including the ideals of equal access to communication, institutionalised editorial freedom, and support for cultural media policies and consensual decision-making, all aiming at an inclusive and diverse sphere of public communication (Jakobsson et al., 2021; Syvertsen et al., 2014).

As much of hate speech today is produced online on digital platforms, hate speech has also become a specific media- or platform problem, raising questions about how to allow or limit hate speech, and how the online platforms themselves ought to take part in this effort. As the online sphere is dominated by international platform giants, a risk exists that the platforms are increasingly becoming the dominant definers of the public's communication values that are to be protected under their regulation (van Dijck et al., 2018), while states are losing their power to influence the values supported in public communication. Recently, the growing power of online platforms has been addressed by the European Union's Digital Services Act (European Commission, 2020), with a new set of requirements to control online issues such as hate speech and disinformation. Nordic countries have now also joined forces, and through the Nordic Council of Ministers, initiated a think tank that created recommendations to support Nordic democratic debate in the age of Big Tech: debate that is free, inclusive, and free from hate speech and disinformation (Nordic Council of Ministers, 2023).

These new developments show that especially online hate speech poses states with a new set of challenges, asking for greater clarity on a policy level, and therefore it is no wonder that already before the newest international efforts it has been an issue of Nordic (media) policies in recent years. Traditionally, strong media policies have been a central element of steering and governing the Nordic media welfare states. When faced with new challenges, the Nordic countries have been keen to use the tool of media policy – and hate speech has been no exception.

The Nordic research interest on these hate speech policies has until now been mainly of a legal nature (e.g., Bladini, 2017; Viljanen, 2019). In this chapter, I approach the policy papers from a wider perspective by analysing how the current hate speech issue and related challenges are perceived in the

Nordic countries. The purpose of the chapter is explorative and descriptive, and the main research questions are as follows:

- RQ. How have the Nordic states defined and addressed the challenge of hate speech on the policy level, and how could the principles of the media welfare state help or not help to deal with hate speech in the current digital media environment?

With these general research questions in mind, and following the hate speech conceptualisations by Brown (2017a, 2017b), I more specifically analyse Nordic hate speech policy documents ($N = 11$) with a focus on the possibly harmful functions of hate speech defined in the policies, the perceived victims of hate speech, what rights or values are perceived as possibly threatened and how, and finally, the suggested remedies or actions for the perceived hate speech issues. Special attention in the analysis is paid to the role of the online media sphere as the forum of hate speech and related solutions for the protection of public and communication values or rights in an age of digital media and platformisation, within the specific context of the Nordic media welfare states.

The chapter begins with discussion on how hate speech has been defined and experienced in Europe and the potential harms it creates, especially in the Nordic and digital contexts. This is followed by an analysis of the recent hate speech policy documents, and finally conclusions and a discussion. The chapter ends concluding that when being faced with new challenges such as online hate speech, the Nordic policies rely on remedies that are familiar to media welfare states: support for free and independent media as well as diversity of information, and striving towards equal possibilities for well-informed and media-literate citizens through digital media. International regulation and joint Nordic actions towards the platform giants, on the other hand, can offer concrete solutions for the future.

Defining hate speech

“Hate speech” lacks one standard international definition, but the general idea of the type of utterances the term refers to is well outlined by the following conceptualisation:

Hate speech is to be understood [...] as the advocacy, promotion or incitement, in any form, of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization or threat in respect of such a person or group of persons and the justification of all the preceding types of expression, on the ground of ‘race’, colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status. (ECRI, 2015)

The above definition also illustrates the broad scope of targets and acts of hate speech. In Europe, the issue of hate speech has been persistent, and migrant, racial, ethnic, and religious minorities have been particularly targeted (e.g., Arcila-Calderón et al., 2022). The so-called European refugee crisis in 2015 caused a peak in hate speech directed at asylum seekers, which also affected other migrant populations in Europe (Berry et al., 2015; Greussing & Boomgaarden, 2017; Pöyhtäri et al., 2019). This, as well as the Islamist terror attacks in several European countries from 2004 onwards, have incited hatred towards Muslim populations in general, adding to the political discussion on failed integration and multiculturalism (Vertovec & Wessendorf, 2010). The discussion has especially been fuelled by many populist and nationalist leaders in Europe, openly using anti-migrant sentiments and calling for restrictive migration policies (Tumber & Waisbord, 2020). In addition, many European countries have seen a rise in antisemitism and far-right movements (see, e.g., FRA, 2016.) Lately, the war on Ukraine has increased anti-Russian sentiments (Saresma et al., 2022).

Online misogyny has become a growing issue, targeting especially women in politics, journalism, research, and other public positions, but also women in general (e.g., Amnesty International, 2018; Article 19, 2020; Posetti et al., 2020; Van Sant et al., 2021). Similarly, gender- and sexual minorities have been reported to be severely attacked online. In an era of growing populism, conspiracy theories and “post-truth”, hate speech and online harassment are increasingly used to assault those in public and powerful positions, with the purpose of undermining their legitimacy (Tumber & Waisbord, 2020), hindering their work, or silencing them, which has also been a growing concern in the Nordic countries. For example, police officials, judges, other officials, politicians, journalists, researchers, and civil society activists have reported on increased online harassment, leading to increased support for these groups and inquiries into new legislation initiatives (related to Finland, e.g., Illman, 2020; Mäkinen, 2019).

In the summer of 2023, a heated debate on racism, hate speech, and online violence in Finland took place as some old, overtly racist social media messages of recently elected parliament members and ministers of the radical right Finns Party became public. The following public debate also demonstrated cases of coordinated online harassment of those who were critical towards the politicians. As a result, the government needed to publish an official declaration on equality, racism, and hate speech in August 2023. Sweden and Denmark have been under terrorist threat due to heated societal debate on freedom of expression and blasphemy, and the governments have needed to clarify their position in international fora. These examples show that the issue of hate speech is high on the public agenda and therefore also asks for political and policy attention.

As such, hate speech is considered an endangerment to the cohesion of democratic societies, the protection of human rights, and the rule of law, and it possibly enhances the risk of societal unrest and violence (see Council of Europe, 2023). In its most severe forms, hate speech violates the principles of international treaties such as the Universal Declaration of Human Rights (UDHR, 1948) and the International Covenant on Civil and Political Rights (United Nations, 1966). These treaties recognise the rights of each individual to, for example, equality, personal dignity, security, and freedom of opinion and expression (UDHR, 1948: articles 1–3, 19), and they forbid all forms of discrimination in violation of those rights (UDHR, 1948: article 7). These rights need to be weighed against each other, and the right for freedom of expression does not allow an attack on other rights. These principles are also reflected in national legislations of the Nordic countries.

Within the Council of Europe, the European Court of Human Rights (www.echr.coe.int) monitors the application of the European Convention on Human Rights, including issues pertaining to freedom of speech and hate speech. Despite the described concerns on hate speech, protection of freedom of speech is strong within Europe, and convictions on illegal hate speech are generally rare. According to European Court of Human Rights case law, speech that is societally worrying or insulting is to be tolerated. There are, however, several charters and recommendations setting human rights standards and providing guidelines for the Council of Europe and European Union member states (in 2023, the Council has 46 member states and the European Union 27) to deal with hate speech and to support the victims in the European context. Moreover, both the Council and the European Union have dedicated measures in recent years to addressing the issue (see Council of Europe, 2023; European Commission, 2021 on hate speech; see also Pingen, 2022).

Beyond legal definitions, hate speech is more generally a phenomenon of the current, platformed and digital societies. It exists both online and offline, but in today's digital world, we most typically encounter it in the online environment, whether it concerns ourselves or others. In this more everyday sense, hate speech is usually defined broadly, and it contains various types of behaviours online, ranging from violent threats to discriminatory, derogatory, humiliating or otherwise harmful, uncivil, or disturbing speech that is nevertheless not illegal. In a broad definition, hate speech can refer to, for example, verbal discrimination or attacks against various non-ethnic minorities, political hate speech, misogyny, violent pornography, online bullying and harassment, trolling, or doxing – and it has also been referred to, for example, as cyberhate (Brown, 2018; Edelstein & Wolf, 2013), cyber violence (United Nations Broadband Commission, 2015), toxic speech (e.g., Amnesty International, 2018), or extreme speech (Pohjonen & Udupa, 2017).

As an online phenomenon, hate speech has become very mundane, and as such, difficult to avoid.

Altogether, this broad scope of behaviours and utterances possibly classifiable as hate speech complicates the everyday understanding of, or chance to reach, a general consensus on what constitutes hate speech. An unsettled debate concerns how we are to understand hate speech, what potentially can be regarded as an act severe enough to actually constitute (illegal) hate speech, which groups should be protected from hate speech, and whether in order for something to constitute hate speech, the harms caused should always be actual and direct, excluding societal and indirect harms (e.g., Article 19, 2019; Brown, 2017a, 2017b; Calvert, 1997; Udupa & Pohjonen, 2019).

Those advocating the most liberal viewpoints often claim that any restrictions on freedom of expression based on hate speech violate the basic right of free speech (e.g., Brown 2017a; Citron, 2014; Molnar & Malik, 2012). Differences in the perceptions are especially clear when comparing the European and American views on the matter, with the US representing a very liberal view and the European perspective stressing the ideals of the common good in addition to the freedom of expression. These differences have also been reflected in the debate on the responsibilities of online platforms (see, e.g., Horowitz & Nieminen, Chapter 4).

Theoretically, hate speech has also been approached as a discourse, a form of othering that does not necessitate that actual or overt hatred is expressed in words. Speech act or discourse can contain a covert expression of hatred, embedded in the specific context of the speech act, or it can be expressed, for example, by means other than words (e.g., Baider, 2020; Brown, 2017a, 2017b; Laaksonen et al., 2020 Pohjonen & Udupa, 2017; Saresma et al., 2022). Similarly, research on online hate and racism has discussed the covert forms of hate speech, textual, algorithmic and platformed practices that are not necessarily direct expressions of hate but that support the circulation of hate and are used to stir up hatred as well as support hate communities (Anat & Matamoros-Fernandez, 2016; Brown, 2018; Daniels, 2018). Such discourses do not necessarily have concrete, real-life consequences, but they contribute to the overall atmosphere regarding, for example, minorities. They are discourses connected to positions of power; individuals are often the targets of hate as representatives of a minority group (e.g., Brown, 2017a), but as we see today, targets can also be those perceived as powerful.

Discursively, hate speech has also been referred to as an umbrella or family resemblances concept, as it describes a set of phenomena that do all have a common denominator in some sense, yet are not exactly the same (Brown, 2017b). For the purposes of this chapter, it is useful to approach hate speech as such a broad, family resemblance concept that the relevant policy documents define in more detail (this approach is further specified under the Research material and methods section).

Hate speech thus remains an issue of disputed definitions and varied vocabulary. Beyond this, hate speech is also context-dependent, and as such, sensitive to societal and other factors (Brown, 2017a). Because of the contested and contextual character of hate speech, it is relevant to have a closer look at the recent Nordic policy definitions concerning the phenomenon to better understand what is at stake in the current Nordic hate speech debate and policies.

Hate speech in the Nordic countries

As in most other countries, “hate speech” as a concept is not defined in Nordic legislation. However, in all Nordic countries, certain verbal acts containing speech that can be categorised as hate speech have been criminalised. Such acts include, for example, defamation, illegal threats, persecution, and incitement to hatred. For the latter, the protected characteristics in Nordic legislations include race, skin colour, ethnicity, religion, nationality, and sexual orientation. The legislations vary in terms of other protected characteristics, but, for example, gender is in general not a protected characteristic – a deficiency that is now under scrutiny in all Nordic countries. In addition, basically all crimes can be investigated and convicted as hate crimes, if a hate-based motive can be demonstrated (see further on legislation, e.g., Bladini, 2017).

Besides possibly constituting a criminal act, hate speech can also be prohibited by other legislation, as is the case in, for example, Finland, where the 2004 Non-discrimination Act (Ministry of Employment and the Economy, Finland, 2010) and also the Act on Equality between Women and Men (Ministry of Social Affairs and Health, Finland, 1986) define the type of discrimination and harassment that are illegal. Harassment is hereby understood as deliberate or de facto infringement of the dignity of a person based on a specific characteristic of that person, if a degrading or humiliating, intimidating, hostile, or offensive environment towards the person is created by that behaviour. This type of harassment can contain hate speech.

The Nordic countries – as signatories of the above-mentioned international treaties, and as members of the Council of Europe, the European Union, or both – are also bound to these charters and recommendations. Next to that, they also have shared cooperation in issues related to hate speech, especially through the fora of the Nordic Council and Nordic Council of Ministers, the formal bodies for inter-parliamentary and inter-governmental cooperation in the Nordic region. Recently, the Nordic cooperative bodies have discussed the issue of gender-based online hatred, threats, and harassment, as well as other types of cyber threats in times of the Covid-19 pandemic (Nordic Council, 2023). Especially in the area of (gender) equality, the cooperation has been ongoing and strong in the past years, with various reports and recommenda-

tions being published to discuss the problems of online hate and misogyny and to compare the related legislations and measures (e.g., Bladini, 2017; Dragsted Mutengwa et al., 2015; Mogensen & Holding Rand, 2020; Nikulainen, 2021). New recommendations for Nordic democratic debate on digital platforms, discussing disinformation and hate speech specifically, were launched by a working group in early 2023 (Nordic Council of Ministers, 2023).

Nordicom published a special issue of *Nordicom Information* in 2015, compiling current views from Nordic media (and other) research on the growing problem of hate speech and online harassment (Wadbring & Mølster, 2015). Lately, Nordic research on hate speech has also been approached with relevance to social media, disinformation, polarisation, the far-right, populism, and other online phenomena (see, e.g., Horsti & Saresma, 2020; Nikunen et al., 2021; Udupa et al., 2021).

In the Nordic context, hate speech is not only a threat to universal rights, it also endangers the normative preconditions of democratic culture, egalitarianism, and social inclusion that are inherent to the Nordic model. This model requires a well-functioning public sphere of open social and political deliberation (Engelstadt et al., 2017), enabled by equality, open society, and access to information (see Carlsson & Pöyhtäri, 2017). These values have also been supported by the ideal of the media welfare state through equal access to communication, institutionalised editorial freedom, supportive cultural media policies, and consensual decision-making, all aiming to support an inclusive and diverse sphere of public communication (Jakobsson et al., 2021; Syvertsen et al., 2014).

While hate speech on the one hand poses a challenge to this value system, which is perceived as specifically Nordic, the value system as such could on the other hand offer tools to resist the observed harms of hate speech. Therefore, in this chapter, I also look at how the suggested actions to address hate speech relate to the media welfare state ideals, or whether some new modes of action outside the media welfare state thinking are emerging.

Hate speech as a threat to rights and public values on online platforms

In the era of digital platforms and social media, hate speech is especially associated with the online environment (e.g., Alkiviadou, 2018; Brown, 2018; see also Horowitz & Nieminen, Chapter 4). Hate speech is not platform-specific in the sense that it can be found and observed on all types of online platforms, ranging from the comment sections of online newspapers to message boards, small online discussion fora, and the various social media platform giants, such as Facebook, Twitter (now known as X), Instagram, or YouTube. The users of various platforms are handy in finding and using

exactly those platforms and affordances related to hate speech that enable their planned actions the best, and therefore the manifestations of hate speech also vary per platform (see, e.g., Bucher & Helmond, 2018; Siaperá & Viejo-Otero, 2021). Additionally, the responsibilities, capabilities, and willingness of larger and smaller platforms to deal with hate speech vary.

With their platform policies, terms of use, and user guidelines, as well as the moderation practices performed, the platforms direct the actions that are allowed. In their platform policies, the platforms are *de facto* acting based on civil law and their privately created contracts between them and their clients, the platform users (see, e.g., Neuvonen, 2021). Experience shows that the contracts that the users sign with the platforms are anything but clear or transparent.

Next to these user contracts, platforms are also regulated by European and national legislations, which I do not discuss in all detail here. Article 14 of the European E-Commerce Directive (European Union, 2000), however, states that digital or online platforms are not legally responsible for hosting illegal content, although they are required to remove such material once it has been flagged, but this obligation only applies to specified content. In 2016, a European Code of Conduct on countering illegal hate speech online was signed by the European Commission and four major IT companies (Facebook, Microsoft, Twitter, and YouTube) in order to respond (on voluntary basis) to the proliferation of racist and xenophobic hate speech specifically online (European Commission, 2016). Later on, other Internet giants have joined the conduct, giving them responsibilities to remove illegal flagged content from their platforms. In the monitoring exercise of late 2021, 81 per cent of the flagged content was assessed by the platforms within 24 hours, and 62.5 per cent of the content deemed to be illegal hate speech was removed (in comparison, 40% and 28%, respectively, in 2016). Companies' responsibilities for illegal content, as well as more transparency on their actions towards the users, have been under scrutiny in the two major legislative initiatives by the European Commission since 2020, The Digital Services and the Digital Markets Act, now accepted by the European Parliament. These two acts aim for safer online space with a protection of fundamental rights for all users and will also necessitate changes to existing national legislations.

Most important to note for the purposes of this chapter is that, as actors, the platforms today are important guardians and implementers of public values, as pointed out by van Dijck, Poell, and de Waal (2018). "Public value" is the value that an organisation contributes to society to benefit the common good (Moore, 1995). Hereby, platforms are not the only contributors to these values, but in their actions and with the actions that they allow for, or require from, their users, they take part in the creation of the online spheres in which the social interaction takes place. These spheres can support human rights and public values – or they can work against them.

As several revelations and studies have in recent years made clear, the platform giants are currently making greater effort to maximise their financial profits, instead of supporting the public values enhancing the common good and really trying to solve the problems of hate speech and other harmful content (Pelley, 2021). Their algorithms have been found to support the spreading of hateful content and be biased in moderation. Further, next to spending money on lobbying for favourable legislation, the platforms are also making money by allowing advertisements related to hateful content (Dwoskin et al., 2021; Faife, 2021; Hindman et al., 2022; Lomas, 2022; Yin & Sankin, 2021). There is a real risk that by allowing and supporting such practices, the platforms are becoming the strongest definers of commonly accepted norms and values online, or the common good, whereas states are losing their power to stay in control.

Hereby, what is understood as a value or “the common good” is historically and ideologically variable, and they are defined through various collective actions and democratic debates about the foundations of a society. Ideally, this negotiation should be the shared responsibility of all societal actors – the Nordic countries being a well-known example of such collective negotiations, also in relation to the media specifically (Syvertsen et al., 2014). Traditionally, however, and in practice, public institutions have especially been the designated custodians of “the common good”, and the desired public values have been cemented into laws, regulations, and policy papers, with relevance to media (van Dijck et al., 2018).

The studied hate speech policy papers not only define the issue as such, but also the values to be supported and perceived as the common good. They also formulate propositions concerning the desired norms and actions related to hate speech in society, including the platforms.

Research material and methods

The research material for this study consists of Nordic policy documents ($N = 11$) related to hate speech and the digital media environment, dating 2016–2021. The policy papers were found through an online search on the governmental websites of Sweden (regeringen.se), Norway (regjeringen.no), Denmark (regeringen.dk), and the websites of various Finnish ministries between November 2020–January 2021 and April–June 2022. The search terms related to hate speech are further specified in Table 3.1. The studied policy documents were selected on the basis of relevance: First, hate-speech *specific* documents were included; if not available, those policy papers that discussed the related areas of free expression, media, or online-related matters encompassing hate speech were included. Also, some policy papers discussing hate speech faced by specific minorities were included in the selection, if these policies also related to matters of free expression and digital media.

This choice was made to keep the focus of the analysis on the current media environment, and not expand it too much into other areas of society (e.g., discrimination in school or workplace environments, or sexual harassment). The selection of documents affects the findings, of course, but the central research questions can be covered best with the selected data.

The method of study can be described as analysis of policy papers, in which the term “policy” is broadly understood as “a course of action (or non-action) taken by a government or legislature with regard to a particular issue” (Knill & Tosun, 2012: 4). Policies can be implemented through legislative processes, leading to the passage of ordinances. Hate speech policies can also be, and often are, implemented through various types of executive or agency actions, or through the work of independent commissions (Wolman, 2019). In this case, I was, however, not studying the actual policy actions taken by governments or other actors, but instead was interested in how the hate speech issue itself, and the desired course of action on the topic, are defined in the relevant policy papers.

The method of text analysis was inspired by discourse analysis, and I was interested in how the concept of hate speech was used, especially in terms of the harms it is considered to be causing. According to Brown (2017b: 567–573), potential functions that the concept of hate speech relates to in various discourses include at least, but are not limited to, the following:

- (1) highlighting forms of harmful speech [especially targeted at specific underprivileged groups in society], (2) flagging up socially divisive forms of speech [highlighting social cleavages in socially diverse societies], (3) identifying forms of speech that can undermine people’s sense of equality [e.g., through inciting discrimination], (4) articulating civility norms [in public speech], and (5) labelling up forms of speech that undermine democracy [causing threats to essential values in societies]. (Brown, 2017b: 573)

These functions are not fully inseparable, and they overlap each other. Yet, they provide a tool to analyse and discuss the discursive and actual variety of hate speech and the potential harms it causes. I track these functions in the policy documents by analysing whether and how they are mentioned in the policies.

The focus is furthermore on who the perceived victims of hate speech are, what rights or basic values in the Nordic context are perceived as threatened, and what the possible solutions to the issues are, especially with regard to the digital or online media environment.

Analysis: Hate speech as a threat to Nordic democratic values

The Nordic hate speech policy documents have been produced by several ministries or areas of governance, ranging from justice, interior issues, and equality issues to education and culture, or as a joint effort of the government (see Table 3.1). This also indicates the complexity of the hate speech issue that asks for a broad approach, but also the variety in placing the issue within the governmental structures in various countries. The policy documents found and analysed are products of the responsible ministries or the result of expert consultations. One of the Finnish documents was produced through citizen participation. Some of the policies discuss hate speech specifically, while others discuss it with relevance to democracy issues, freedom of expression, or IT and media. This also shows the variety of perspectives taken on the topic.

Despite this variance, the Nordic policies in general share an understanding of the hate speech issue. Hate speech is foremost perceived to undermine some of the basic functions and values of the democratic societies, especially freedom of expression and opinion, free public debate, and respect for human rights and equality. There is a shared view that no opinion should be left out in public debate and a great variety of free expression in the debate should be supported, but that the participants should respect each other. This is currently considered to be threatened by hate speech, threats, disinformation, and propaganda.

In all the countries studied, the current hate speech problem is analysed in the context of the digital era and online sphere. The policy documents discuss how the rise of social media and online platforms has not only provided people with new means for self-expression and participation, but also created multiple problems, including hate speech. The type of hate speech that the policy documents discuss as a problem does not refer to the general incivility of the public debate, as this needs to be tolerated and supported as part of free expression, no matter how disturbing it might be (see, e.g., Mäkinen, 2019). Problematic hate speech instead is characterised as truly harmful, partly illegal, and causing a variety of societal harms.

Harmful hate speech is linked with other harmful behaviours, such as threats and harassment. All the documents start their hate speech definitions with references to existing criminal (and other) laws, pointing to hate speech that attacks certain named qualities of persons and groups, thus especially minorities, such as migrant, ethnic, sexual, and religious minorities. These laws recognise the harms that these minorities are facing, yet the hate speech that the policy documents discuss in greater detail refers to public debate and freedom of expression, in which those that take part in this debate through their public position or profession are especially being threatened. Hate

speech is thus perceived as undermining democracy and its functioning in general through causing harms to public debate, democratic participation, and public institutions, even more than it is seen to be harmful to specific minorities (see Brown's hate speech functions 1 and 5 above).

Typical victims of the hate speech that undermines democracy and its functions include politicians, officials, representatives of local councils, journalists, writers, artists, and researchers. The concern for the threatened professionals, free speech, and other democratic values is strongly voiced in the Finnish and Swedish documents that concentrate on this issue, but it is also central in the Norwegian and Danish documents analysed (see, e.g., Ministry of Justice of Denmark, 2020b: 539–570). It can be stated that the focus of the hate speech problem has slightly shifted from purely intolerance-based hatred directed at certain minority groups, as defined, for example, in the above-mentioned hate speech definition of ECRI (2015), towards a more general, or democracy-orientated hate speech definition. This type of hate speech manifests itself in public debates on political or controversial issues, or in matters of public governance, and it mainly targets individuals based on their status or public role.

After the publication of the analysed policies, the Nordic countries have also re-examined their legislation related to online harassment (Finland), hardened the criminal sanctions for harassment (Denmark, Sweden), given more resources to the police to prevent and investigate harassment (Sweden), and created a website to support those in politics and other public functions facing harassment (Norway). Some of the policy papers have also been updated after mid 2022. These actions all indicate the perceived severity of the threats in public debate that hate speech and harassment create.

In addition, the hate speech issue is approached with a broad scope, acknowledging that it has its roots in various societal processes, such as growing differences in shared basic values, polarisation, growing segregation in societies in various areas, less support for diversity through the rise of populism and nationalism, the changing media landscape, and difficulties in the sustainability for quality journalism that creates grounds for disinformation. This is where the hate speech definitions in the policies touch upon the functions of supporting discrimination and social divisions, also addressed in Brown's (2017b: 567–573) categorisation of hate speech. Such social divisions are seen as a source of hate speech, as they help maintain societal inequalities and discrimination, but these divisions also manifest themselves through the hate speech uttered.

With reference to these social inequalities and divisions, the policy documents demonstrate a strong discourse on the democratic welfare-state and the idea that no one should be left behind: Everyone deserves to be a part of a functioning democracy with equal possibilities for freedom of expres-

sion and public participation (see especially Government Offices of Sweden, 2020). This thinking is paired with the values of freedom of opinion, free public debate, human rights, diversity, equality, and openness. These need to be protected and guaranteed by the state, in this case by tackling hate speech, but people in general can and should also do their share to make a well-functioning democracy and public debate possible.

It is also expressed that those who produce hate speech are somehow unable to function in the society: They have fallen victim to the current changes in both ordinary society as well as the current information societies, and they need to be assisted in finding the right path. For example, extreme thoughts can be prevented with social policies and education, working against segregation in society and through offering alternatives to youth (Norwegian Ministry of Children and Equality, 2015; Government Offices of Sweden, 2020).

Acknowledging the wide-reaching causes and effects of hate speech in society, the suggested remedies to hate speech also have a wide coverage, involving governance, officials, legislation, education, social policies, the perpetrators, the victims, users of online services, the media, and finally the online platforms. As the Swedish policies suggest, protecting a democratic society is everyone's responsibility. Central to this goal is to strengthen democracy through strengthening people's possibilities to resist disinformation, propaganda, and online hate, for example, through improved media and information skills (Government Offices of Sweden, 2020: 168–169), as well as through offering places for constructive meetings and debates (Norwegian Ministry of Children and Equality, 2015). This also protects society and the weakest groups from polarisation and extremism, as is stated in the Norwegian strategy against hate speech.

When it comes to expectations expressed towards the big international tech companies and online platforms, the policies do not appear very determined. The policies mention the general and ongoing developments within the European Union, especially the Digital Services Act, which was approved by the European Parliament in early 2022. The Digital Services Act aims to create stronger obligations and clearer guidelines for the governance of the practices and curation of content on online platforms, especially the largest ones. Much, however, will depend on the national implementations of the act in the coming years.

Beyond these European processes, there does not seem to be much hope or expectations for quick or definite solutions concerning platform regulation. The value most commonly called for in the policies is transparency towards the users of the platforms on the principles of online content moderation (the notice and take-down procedures defined, e.g., in the e-Commerce Directive). For example, compulsory user registration with real identity does not receive support, nor does extending the editorial responsibilities of large interna-

tional social media platforms (e.g., Ministry of Justice of Denmark, 2020b: 656–663). However, the fact that the platforms decide about and steer the democratic debate, information flows, and uses of user data is understood as a problem for democracy in itself. In Denmark, a specific policy document dedicated to making Big Tech companies more accountable and democratic was published in 2021 (Danish Government, 2021), but it did not offer any practical solutions. The same worry of platforms steering the course of democracy is expressed in the recommendations of the Nordic think tank on democratic debate (Nordic Council of Ministers, 2023).

More hope regarding the hate speech issue is put on the national media and platforms by discussing a need to clarify their position and responsibilities as online media through bringing the existing media laws up to date with the current digital media environment, and possibly stretching the responsibilities not only to national news media, but also to various types of national online media and platforms by ensuring that the laws throughout are technology-neutral (as suggested in Norway). All the hate speech policies analysed indicate the central role of free media in society and ask for more finances and other types of support for their work in order to enhance the public's access to free and objective information. Debates have also been ongoing with national online platforms, but a willingness to participate in the negotiations has not always translated itself into changes in the actual policies or practices of the platforms (as in Finland). As the policies seem somewhat powerless towards the private online companies or platforms, all policies and most suggested remedies involve governance measures, eventual legislation initiatives, civil society actions, and online users and their education.

The description of the hate speech issue per country is further summarised in Table 3.1.

Table 3.1 Hate speech issue in Nordic policy documents

	Finland	Sweden	Norway	Denmark
Documents	Ministry of Interior: Expert working group report: <i>Words are actions. More effective measures against hate speech and cyberbullying</i> (Mäkinen, 2019) Ministry of Justice: Recommendations of the "Against hate" project (Aaltonen, 2019) Ministry of Finance & Ministry of Justice: <i>Kansalaispaneeli sananvapaudesta. Kansalaispaneelin loppuraportti</i> [Citizen panel on freedom of speech. Final report of the citizens' panel] (Jäske et al., 2021)	Action plan: <i>Till det fria ordets försvar</i> [For free words defence] (Government Offices of Sweden, 2017) Report: <i>Det demokratiska samtalet i en digital tid</i> [The democratic debate in digital times] (Government Offices of Sweden, 2020) <i>Samlat grepp mot racism och hatbrott</i> [Joint grip against racism and discrimination] (Government Offices of Sweden, 2016)	<i>Strategi mot hatefulle ytringer</i> [The Government's Strategy against Hate Speech] 2016–2020 (Norwegian Ministry of Children and Equality, 2015) <i>The Norwegian government's action plan against racism and discrimination on the grounds of ethnicity and religion 2020–2023</i> (Norwegian Ministries, 2019)	<i>Ytringsfrihedens rammer og vilkår i Danmark, del I & II</i> [The framework and conditions of freedom of expression in Denmark, I & II] (Ministry of Justice of Denmark, 2020a–b) White paper: <i>Mod et bedre samfund med tech-giganter</i> [Towards a better society with tech-giants] (Danish Government, 2021)
Hate speech vocabulary (translated)	hate speech, nethate, online hate, assaults, crowd-sourced/organised attacks/dog whistling	hate crime, hate, threat, hate and threat, bullying, net bullying, nethate, insults, harassment, hate propaganda, hate rhetoric	hate(ful) expressions, hate talk, hate, incitement, net incitement, nethate, threats, offences	hate(ful) expressions, hate(ful) talk, illegal expression, hate, threats, harassment, online harassment, insults, hate offences
Functions of hate speech (Brown 2017b): What does hate speech relate to or threaten?	democratic values; harmful speech (minorities); discrimination; social division; civility norms	democratic values; harmful speech (minorities); discrimination; social division; civility norms	democratic values; harmful speech (minorities); discrimination; social division; civility norms	democratic values; harmful speech (minorities); discrimination; social division; civility norms
Who is the target or victim?	those in public positions (justice, police, journalists, politicians, researchers), minorities and other marginalised groups	those in public positions: especially journalists, people in public office (or position of trust), artists, researchers, civil society actors (also: youth, minorities)	especially minorities, but also those participating public debate	opinion leaders (politicians, journalists, artists), women, ethnic minorities (Muslims, Jews), sexual and gender minorities
What are the suggested measures towards tech companies and others?	Discussion and collaboration with the tech giants, especially through the European Union. Collaboration with smaller, national platforms. More broadly, education of police, justice, and a wide range of actors on media literacy, infrastructures of reporting and investigating hate speech cases, more support for victims.	Discussion with the tech giants, especially through the European Union. More broadly, social-policy, education (media literacy); involving police, justice, wide range of actors.	Re-analysing and modifying the existing media laws to better explicate the responsibilities in various types of media (incl. social media) on hate speech. More broadly, support for freedom of expression and equality; support for idea that work against hate speech, discrimination and extremism go hand in hand.	Suggestions for hardened responsibilities for users of illegal content; more transparency required from the online platforms in line with general European Union principles. More broadly, suggestions related to legislation and a wide range of actors.

Conclusion and discussion: Hate speech and the revival of the media welfare state?

Following the categories of hate speech according to Brown (2017b), hate speech manifests itself in the Nordic context as harmful speech directed at minorities, in acts that undermine democracy and its values, as well as a factor that creates a risk for further social discrimination and inequality. Instead, the civility norms of public debate are less addressed in the Nordic policy documents. The general incivility of public online debate is highlighted, but as such, it is not an issue that the Nordic policies aim to erase. On the contrary, freedom of speech in all its forms is protected and valued in the Nordic policies. It is only when the incivility takes on more serious forms that lead to societal harms, such as threats and harassment that silence debate, create inequality, or damage the democratic functions (such as people's right to take part in politics), that the hate speech policies call for action and restrictions on hate speech. Unlike how certain (often conservative or populist) voices in the debate claim, hate speech policies do not advocate for censorship or denial of critical debate; instead, they are an attempt to support freedom of speech and democratic participation rights for everyone in the society.

Altogether, the hate speech functions identified by Brown (2017b) seem to cover the hate speech issue in the Nordic policies well. The discursive and comparative approach on hate speech's functions also sheds light on the multiplicity of hate speeches, causing concern today. The focus on the hate speech issue in the Nordic digital media spheres no longer concerns only protected minorities, but also public functions and the people participating in the public debate. This shift can also be seen in the hate speech vocabulary used (see Table 3.1), which does not only discuss hate speech, but also other types of threats, hateful or disturbing expressions, assaults, organised attacks, and online hatred or bullying. Legally speaking, hate speech is no longer necessarily a group-targeted criminal act (such as incitement to hatred), but an individually targeted act (such as illegal threat or libel directed at a single person).

Discursively, it has been suggested that instead of discussing hate speech, other terms that better address the broad range of hate speech, its contextual-ity, and the variety of uses should be used. For example, the concept "extreme speech" (Udupa et al., 2021) contains two analytical parts: first, *exclusionary* extreme speech that is directed at vulnerable and disadvantaged groups in society to denote social inequalities and to reinforce them; and second, *derogatory* extreme speech, which refers to expressions that in a specific context are, for example, uncivil, provocative, and offensive and can be directed at both minorities and groups holding power, as well as institutions. The terms thus grasp both sides of the current hate speech debate, including both vulnerable and powerful targets. Furthermore, Udupa and colleagues (2021) have

distinguished *dangerous* speech that has reasonable risk to catalyse or trigger actual harm and violence. It can be debated which terms are the best to use, but based on this study, it seems highly necessary to define in each context, and for each separate policy or other action, the precise type of hate speech that is in question and whom it concerns.

With reference to Nordic values and their protection, it is obvious from the policy documents that hate speech forms a challenge that the Nordic countries are not (yet) equipped to tackle. Hate speech is a threat to democracy that is created especially by the digital media environment and platforms, which the Nordic societies cannot alone hold accountable for their actions. Also, the fact that the digital giants have so much power is outlined as a threat to democracy. In this matter, the Nordic community needs to rely on international cooperation and action. What the Nordic societies can do individually and jointly is try to address the challenges with the tools that have always been used: by creating media policies in support of the values of democratic societies, and by providing tools in various areas of society for its continued support.

All Nordic policy documents rely not only on the principles and values of democracy, but also on the elements of the Nordic media welfare states. States are seen to carry the main responsibility for the maintenance of a free, accessible, equal, and safe public sphere, and to support institutions, professional media, and individuals in this joint effort. The policy documents call for continued media support for independent media, diversity of information, and equal possibilities for well-informed and media-literate citizens through digital media. These and other, more concrete measures are also recently proposed in the outcome document of the Nordic think tank on democratic debate in the age of Big Tech (Nordic Council of Ministers, 2023). It expresses that if the Nordic countries want to steer the course of democracy now and in the future instead of letting the big tech or digital platforms do it on our behalf, the countries need to act jointly, based on the media welfare values that have worked well in the past.

The media welfare state ideal may be in crisis, but in such times when it is difficult to keep pace with the new developments, and new solutions are slow to appear, remedies can be found in what is established and known, namely media policies and other soft measures that do not fix the whole problem but do offer relief. Digital challenges such as hate speech ask, if not for the media welfare state's revival, at least for its reinspection and continuous reinvestments that can steer the path towards a future with more control on the type of platforms, public debate, and democracy we want to support in the Nordics.

Acknowledgements

This research was enabled by University of Jyväskylä, Department of Language and Communication Studies, which funded my post-doctoral research in 2019–2021. A special thanks to my research assistant Eero Valkonen, who helped to collect the research material. The chapter was finalised with support from Tampere University and Academy of Finland funded projects “Communication Rights in the Age of Digital Disruption” (Cordi) (funding decision 3208917/2018) and “Everyday Affective Practices Online: Producing, Experiencing and Managing Hate Speech in the Digital Era” (Hffect) (funding decision 340648/2021). The results of this study have been discussed and used as a case example in a book chapter on epistemic rights in the Nordic countries (Pöyhtäri et al., 2024).

References

- Aaltonen, M. (2019). *Against hate – hankkeen suosituksia viharikosten ja vihapuheen vastaiseen työhön* [Against hate – project recommendations for work against hate crimes and hate speech]. Ministry of Justice (Finland). <https://tinyurl.com/yrewmx3u>
- Alkiviadou, N. (2018). Hate speech on social media networks: Towards a regulatory framework? *Information & Communications Technology Law*, 28(1), 19–35. <https://doi.org/10.1080/13600834.2018.1494417>
- Amnesty International. (2018, March 21). *Toxic Twitter – A Toxic place for women*. Amnesty International. <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/>
- Anat, B.-D., & Matamoros-Fernandez, A. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 10, 1167–1193. <https://ijoc.org/index.php/ijoc/article/view/3697>
- Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J.-J., & Blanco-Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. *Comunicar*, 30(71), 21–35. <https://doi.org/10.3916/C71-2022-02>
- Article 19. (2019, December 23). ‘Hate speech’ explained: A toolkit. <https://www.article19.org/resources/hate-speech-explained-a-toolkit/>
- Article 19. (2020). *Online harassment and abuse against women journalists and major social media platforms*. <https://www.article19.org/wp-content/uploads/2020/11/Gender-Paper-Brief-2.pdf>
- Baider, F. H. (2020). Pragmatics lost in defining online hate speech? Overview, synthesis and proposition in defining online hate speech. *Pragmatics and Society*, 11(2), 196–218. <https://doi.org/10.1075/ps.20004.bai>
- Berry, M., Garcia-Blanco, I., & Moore K. (2015). *Press coverage of the refugee and migrant crisis in the EU: A content analysis of five European countries*. UNHCR. www.unhcr.org/protection/operations/56bb369c9/press-coverage-refugee-migrantcrisis-eu-content-analysis-five-european.html
- Bladini, M. (2017). *Hat och hot på nätet. En kartläggning av den rättsliga regleringen i Norden från ett jämställdhetsperspektiv* [Hate and threat on the net: A mapping of legal regulation in Nordic countries from equality perspective]. NIKK – Nordic Information on Gender. https://www.nikk.no/wp-content/uploads/Hat-och-hot-på-nätet_170627.pdf
- Brown, A. (2017a). What is hate speech? Part 1: The myth of hate. *Law and Philosophy*, 36(4), 419–468. <https://doi.org/10.1007/s10982-017-9297-1>
- Brown, A. (2017b). What is hate speech? Part 2: Family resemblances. *Law and Philosophy*, 36(5), 561–613. <https://doi.org/10.1007/s10982-017-9300-x>
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18, 297–326. <https://doi.org/10.1177/1468796817709846>
- Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The Sage handbook of social media* (pp. 233–253). Sage. <https://doi.org/10.4135/9781473984066>
- Calvert, C. (1997). Hate speech and its harms: A communication theory perspective. *Journal of Communication*, 47(1), 4–19. <https://doi.org/10.1111/j.1460-2466.1997.tb02690.x>
- Carlsson, U., & Pöyhtäri, R. (2017). Words of introduction. In U. Carlsson, & R. Pöyhtäri (Eds.), *The assault on journalism: Building knowledge to protect freedom of expression* (pp. 11–27). Nordicom, University of Gothenburg. <https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-5068>
- Citron, D. K. (2014). *Hate crimes in cyberspace*. Harvard University Press.
- Council of Europe. (2023). *Hate speech*. <https://www.coe.int/en/web/freedom-expression/hate-speech>
- Daniels, J. (2018). The algorithmic rise of the ‘alt-right.’ *Contexts*, 17(1), 60–65. <https://doi.org/10.1177/1536504218766547>
- Danish Government. (2021, June 18). *Mod et bedre samfund med tech-giganter. Regeringen, Hvidbok juni 2021* [Towards a better society with tech-giants. White paper June 2021]. <https://www.regeringen.dk/nyheder/2021/regeringen-vil-diskutere-techgiganternes-rolle-med-ny-hvidbog/>

- Dragsted Mutengwa, B., Dehlholm, M., & Nørgaard Christensen, A. (2015, November 2–3). *Hadytringer og sexism i Norden [Hate expressions and sexism in the Nordics]*. ALS Research. <https://www.nikk.no/wp-content/uploads/160309-Opsamlingsnotat.pdf>
- Dwoskin, E., Tiku, N., & Timberg, C. (2021, November 11). Facebook’s race-blind practices around hate speech came at the expense of Black users, new documents show. *Washington Post*.
<https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/>
- ECRI (European Commission against Racism and Intolerance). (2015). *General Policy Recommendation 15 on combating hate speech*.
<https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01>
- Edelstein, Y., & Wolf, C. (2013). *Report and recommendations of Yuli Edelstein, speaker of the Knesset, and Christopher Wolf, chair, National Civil Rights Committee of the Anti-Defamation League, co-chairs of the task force on internet hate of the inter-parliamentary coalition for combating anti-semitism*. ICCA.
<http://hdl.handle.net/20.500.12389/21733http://data.europa.eu/eli/dir/2000/31/oj>
- Engelstad, F., Larsen, H., & Rogstad, J. (2017). The public sphere in the Nordic model. In F. Engelstad, H. Larsen, J. Rogstad, K. Steen-Johnsen, D. Polkowska, A. S. Dauber-Griffin, & A. Leverton (Eds.), *Institutional change in the public sphere: Views on the Nordic model* (1st ed.) (pp. 46–70). De Gruyter. <http://www.jstor.org/stable/j.ctvbkk05k.6>
- European Commission. (2016). *The EU Code of conduct on countering illegal hate speech online*. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en
- European Commission. (2020). *The digital services act package*.
<https://ec.europa.eu/digital-single-market/en/digital-services-act-package>
- European Commission. (2021). *Communication from the Commission to the European Parliament and the Council. A more inclusive and protective Europe: Extending the list of EU crimes to hate speech and hate crime*.
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0777>
- European Union. (2000). *e-commerce directive*. <http://data.europa.eu/eli/dir/2000/31/oj>
- Faife, C. (2021, November 18). Facebook isn’t telling you how popular right-wing content is on the platform. *The Markup*.
<https://themarkup.org/citizen-browser/2021/11/18/facebook-isnt-telling-you-how-popular-right-wing-content-is-on-the-platform>
- FRA (European Union Agency for Fundamental Rights). (2016, November). *Current migration situation in the EU: Hate crime*.
https://fra.europa.eu/sites/default/files/fra_uploads/fra-2016-november-monthly-focus-hate-crime_en.pdf
- Greussing, E., & Boomgaarden, H. G. (2017). Shifting the refugee narrative? An automated frame analysis of Europe’s 2015 refugee crisis. *Journal of Ethnic and Migration Studies*, 43(11), 1749–1774.
<https://doi.org/10.1080/1369183X.2017.1282813>
- Government Offices of Sweden. (2016). *Samlat grepp mot rasism och hatbrott: Nationell plan mot rasism, liknande former av fientlighet och hatbrott [Joint grip against racism and discrimination: National plan against racism, similar forms of hostility and hate crime]*.
<https://regeringen.se/regeringens-politik/samlat-grepp-mot-rasism-och-hatbrott/>
- Government Offices of Sweden. (2017). *Till det fria ordets försvar: Åtgärder mot utsatthet för hot och hat bland journalister, förtroendevalda och konstnärer [For free words defence: Measures against risks of threats and hate among journalists, councillors and artists]*.
<https://www.regeringen.se/information/material/2017/07/regeringen-antar-handlingsplan-mot-hat-och-hot-mot-journalister-fortroendevalda-och-konstnarer/>
- Government Offices of Sweden. (2020). *Det demokratiska samtalet i en digital tid: Så stärker vi motståndskraften mot desinformation, propaganda och näthat [The democratic debate in digital times: So we strengthen resistance against disinformation, propaganda and net hate]*.
<https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2020/09/sou-202056/>

- Hindman, M., Lubin, N., & Davis, T. (2022, February 10). Facebook has a superuser-supremacy problem. *The Atlantic*.
<https://www.theatlantic.com/technology/archive/2022/02/facebook-hate-speech-misinformation-superusers/621617/>
- Horsti, K., & Saresma, T. (2020). The role of social media in the rise of right-wing populism in Finland. In H. Tumber, & S. Waisbord (Eds.), *The Routledge companion to media misinformation and populism* (pp. 376–385). Routledge. <https://doi.org/10.4324/9781003004431>
- Illman, M. (2020). *Järjestelmällinen häirintä ja maalittaminen: lainsäädännön arviointia* [Systematic harassment and dog whistling: Assessment of legislation].
<http://urn.fi/URN:NBN:fi-fe2020120399312>
- Jakobsson, P., Lindell, J., & Stiernstedt, F. (2021). A neoliberal media welfare state? The Swedish media system in transformation. *Javnost – The Public*, 28(1), 1–16.
<https://doi.org/10.1080/13183222.2021.1969506>
- Jäske, M., Sormunen, H., Schauman, J., Sirén, R., Weckman, A., & Rapeli, L. (2021). *Kansalaispaneeli sananvapaudesta. Kansalaispaneelin loppuraportti* [Citizen panel on freedom of speech. Final report of the citizens' panel]. Valtiovarainministeriö, Oikeusministeriö, Åbo Akademi.
https://avoinhallinto.fi/assets/files/2021/03/VM_Kansalaispaneeli_sananvapaudesta_Loppuraportti.pdf
- Knill, C., & Tosun, J. (2012). *Public policy: A new introduction*. Palgrave Macmillan.
- Laaksonen, S.-M., Haapoja, J., Kinnunen, T., Nelimarkka, M., & Pöyhtäri, R. (2020). The datafication of hate: Expectations and challenges in automated hate speech monitoring. *Frontiers in Big Data*, 3(3). <https://doi.org/10.3389/fdata.2020.00003>
- Lomas, N. (2022, 23 April). Report reveals big tech's last minute lobbying to weaken EU rules. *TechCrunch*. <https://techcrunch.com/2022/04/22/google-facebook-apple-eu-lobbying-report/>
- Ministry of Employment and the Economy, Finland. (2010, January). *Non-discrimination act (21/2004)* [Unofficial translation].
<https://www.finlex.fi/fi/laki/kaannokset/2004/en20040021.pdf>
- Ministry of Justice of Denmark. (2020a). *Ytringsfrihedens rammer og vilkår i Danmark: Betænkning afgivet af Ytringsfrihedskommissionen. Del 1*. [The framework and conditions of freedom of expression in Denmark: Report submitted by the Freedom of Expression Commission. Part 1].
https://www.justitsministeriet.dk/sites/default/files/media/Pressemeddelelser/pdf/2020/betaenkning_nr_1573_2020_del_1.pdf
- Ministry of Justice of Denmark. (2020b). *Ytringsfrihedens rammer og vilkår i Danmark: Betænkning afgivet af Ytringsfrihedskommissionen. Del 2*. [The framework and conditions of freedom of expression in Denmark: Report submitted by the Freedom of Expression Commission. Part 2].
https://www.justitsministeriet.dk/sites/default/files/media/Pressemeddelelser/pdf/2020/betaenkning_nr_1573_2020_del_2.pdf
- Ministry of Social Affairs and Health, Finland. (1986). *Act on equality between women and men* [Unofficial translation from Finnish].
https://www.finlex.fi/fi/laki/kaannokset/1986/en19860609_20160915.pdf
- Mogensen, C., & Holding Rand, S. (2020). *The angry internet: A threat to gender equality, democracy and well-being*. Centre for Digital Youth Care.
https://cfdp.dk/wp-content/uploads/2020/11/CFDP_the_angry_internet_ISSUE.pdf
- Molnar, P., & Malik, K. (2012, April 30). *Why hate speech should not be banned*. Free Speech Debate. <http://www.freespeechdebate.com/en/discuss/why-hate-speech-should-not-be-banned>
- Moore, M. (1995). *Creating public value: Strategic management in government*. Harvard University Press.
- Mäkinen, K. (Ed.). (2019). *Words are actions: More efficient measures against hate speech and cyber-bullying*. Publications of the Ministry of the Interior 2019: 30.
<http://urn.fi/URN:ISBN:978-952-324-301-9>
- Neuvonen, R. (2021, January 14). *Yksityiset perusoikeuksien valvojina* [Private companies as guardians of constitutional rights]. Perustuslakiblogi [Constitution blog] [Blog post].
<https://perustuslakiblogi.wordpress.com/2021/01/14/riku-neuvonen-yksityiset-perusoikeuksien-valvojina/>

- Nikulainen, E. (2021). *Preventing online gender-based violence in the Nordics: Key takeaways from the conference “Gender-based hate, threat and harassment on the internet” on 10–11 June 2021*. TemaNord 2021:549. <http://dx.doi.org/10.6027/temanord2021-549>
- Nikunen, K., Hokka, J., & Nelimarkka, M. (2021). Affective practice of soldiering: How sharing images is used to spread extremist and racist ethos on Soldiers of Odin Facebook site. *Television and New Media*, 22(2), 166–185. <https://doi.org/10.1177/1527476420982>
- Nordic Council. (2023). *News*. <https://www.norden.org/en/nyheder>
- Nordic Council of Ministers. (2023). *A Nordic approach to democratic debate in the age of big tech: Recommendations from the Nordic think tank for tech and democracy*. <https://pub.norden.org/nord2023-004/nord2023-004.pdf>
- Norwegian Ministries. (2019). *The Norwegian government’s action plan against racism and discrimination on the grounds of ethnicity and religion 2020–2023*. <https://tinyurl.com/4mmcjwn8>
- Norwegian Ministry of Children and Equality. (2015). *The government’s strategy against hate speech 2016–2020*. <https://tinyurl.com/4f5bstop5>
- Pelley, S. (2021, October 4). Whistleblower: Facebook is misleading the public on progress against hate speech, violence, misinformation. *CBS News*. <https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03/>
- Pingen, A. (2022). *Initiative to extend list of EU crimes to hate speech and hate crime*. EUCrim 4/2021, 221 [Updated January 29, 2022]. <https://eucrim.eu/news/initiative-to-extend-list-of-eu-crimes-to-hate-speech-and-hate-crime/>
- Pohjonen, M., & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. *International Journal of Communication*, 11, 1173–1191. <https://ijoc.org/index.php/ijoc/article/view/5843>
- Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J., & Waisbord, S. (2020). *Online violence against women journalists: A global snapshot of incidence and impacts*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000375136>
- Pöyhtäri, R., Nelimarkka, M., Nikunen, K., Ojala, M., Pantti, M., & Pääkkönen, J. (2019). Refugee debate and networked framing in the hybrid media environment. *International Communication Gazette*, 83(1), 81–102. <https://doi.org/10.1177/1748048519883520>
- Pöyhtäri, R., Neuvonen, R., Ala-Fossi, M., Lehtisaari, K., & Hilden, J. (2024). Nordic illusion and challenges for epistemic rights in the era of digital media. In M. Aslama Horowitz, H. Nieminen, K. Lehtisaari, & A. D’Arma (Eds.), *Epistemic rights in the era of digital disruption* (pp. 171–185). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-45976-4_12
- Saresma, T., Pöyhtäri, R., Knuutila, A., Kosonen, H., Juutinen, M., Haara, P., Tulonen U., Nikunen, K., & Rauta, J. (2022). *Verkkoviha: Vihapuheen tuottajien ja levittäjien verkostot, toimintamuodot ja motiivit* [Nethate: The networks, actions and motives of producers and spreaders of online hate speech]. Valtioneuvoston kanslia [Chancellery of the Government]. <http://urn.fi/URN:ISBN:978-952-383-298-5>
- Siepera, E., & Viejo-Otero, P. (2021). Governing hate: Facebook and digital racism. *Television & New Media*, 22(2), 112–130. <https://doi.org/10.1177/1527476420982232>
- Syvetsen, T., Enli, G., Mjøs, O. J., & Moe, H. (2014). *The media welfare state: Nordic media in the digital era*. University of Michigan Press. <https://doi.org/10.3998/nmw.12367206.0001.001>
- Tumber, H., & Waisbord, S. (Eds.). (2020). *The Routledge companion to media disinformation and populism*. Routledge. <https://doi.org/10.4324/9781003004431>
- UDHR (Universal Declaration of Human Rights). (1948). *United Nations General Assembly Resolution 217A*. <https://www.un.org/en/universal-declaration-human-rights/index.html>
- Udupa, S., Gagliardone, I., & Hervik, P. (Eds.). (2021). *Digital hate: The global conjuncture of extreme speech*. Indiana University Press.
- Udupa, S., & Pohjonen, M. (2019). Extreme speech and global digital cultures. *International Journal of Communication*, 13, 3049–3067. <https://ijoc.org/index.php/ijoc/article/view/9102>
- United Nations. (1966, December 16). *International covenant on civil and political rights*. <https://www.ohchr.org/EN/ProfessionalInterest/Pages/CCPR.aspx>

- United Nations Broadband Commission. (2015). *Cyber violence against women and girls: A world-wide wake-up call* [A Report by the UN Broadband Commission for Digital Development Working Group on Broadband and Gender]. <https://en.unesco.org/sites/default/files/genderreport2015final.pdf>
- van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. <https://doi.org/10.1093/oso/9780190889760.001.0001>
- Van Sant, K., Fredheim, R., & Bergmanis-Korats, G. (2021). *Abuse of power: Coordinated online harassment of Finnish government ministers*. NATO Strategic Communications Centre of Excellence. <https://tinyurl.com/4z2w69h9>
- Vertovec, S., & Wessendorf, S. (Eds.). (2010). *The multiculturalism backlash: European discourses, policies and practices*. Routledge. <https://doi.org/10.4324/9780203867549>
- Viljanen, J. (2019). Combating hate speech online. In M. Susi (Ed.), *Human rights, digital society and the law: A research companion* (pp. 225–238). Routledge. <https://doi.org/10.4324/9781351025386>
- Wadbring, I., & Mølster, R. (2015). Hets och hat online [Incitement and hate online]. *Nordicom Information*, 37(3-4). <https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-4455>
- Wolman, A. (2019). Combating hate speech at the local level: A comparison of East Asian and European approaches. *Nordic Journal of Human Rights*, 37(2), 87–104. <https://doi.org/10.1080/18918131.2019.1614255>
- Yin, L., & Sankin, A. (2021, 8 April). Google has a secret blocklist that hides YouTube hate videos from advertisers – but it’s full of holes. *The Markup*. <https://themarkup.org/google-the-giant/2021/04/08/google-youtube-hate-videos-ad-keywords-blocklist-failures>