Revisiting Digital Policy Through a Feminist Lens

GERGANA **TZVETKOVA**
Research Fellow at Ca' Foscari University of Venice (2022-2024) and co-founder of The Counterintuitive Institute.
E-mail: gergana.tzvetkova@unive.it

ABSTRACT
This article maps and discusses recent key legislative and policy development in combating cyber violence against women and online hate speech, as well as contributions from research, advocacy, and practice linked to the adoption of a feminist approach to digital policy. We review the current policies and recent reports of two big social media platforms – Facebook and TikTok – on how they address these malicious phenomena and propose ways to upgrade these policies by integrating feminist principles.

Questo articolo analizza e discute i principali recenti sviluppi legislativi e politici nel contrasto alla violenza digitale contro le donne e ai discorsi d'odio online, nonché i contributi provenienti dalla ricerca, dall'attivismo e dalla pratica, volti all'adozione di un approccio femminista alle politiche digitali. In particolare, si presterà attenzione alle politiche attuali di due grandi piattaforme di social media – Facebook e TikTok – esaminando come affrontino questi fenomeni e mettendo in luce come l'integrazione di principi femministi potrebbe contribuire a migliorarle.

KEY WORDS
digital policy, feminism, cyber violence against women, social media, hate speech

politica digitale, femminismo, violenza digitale contro le donne, social media, discorsi d'odio

ISSN 1825-0173

Revisiting Digital Policy Through a Feminist Lens

GERGANA TZVETKOVA

1. *Introduction*

The main objective of this paper is to review and map recent key legislative and policy developments taking place at the European Union (EU) level in combating cyber violence against women (CVAW) and online hate speech, as well as contributions from scholars, researchers, civil society organizations, and advocacy experts that discuss the potential benefits of integrating feminist perspectives and adopting a feminist approach to the design and implementation of digital policy on a supranational and national level.

Employing in-depth, systematic literature review, qualitative document analysis and policy analysis, we pinpoint those areas related to the identification of online risks, content moderation and reporting on social media platforms where the application of a feminist and a gender lens may contribute to a digital policy's effectiveness. We review the current policies and recent reports of two big social media platforms, Facebook and TikTok, and discuss how the way they address CVAW and hate speech could be improved by upgrading these policies through the integration of feminist principles. The paper touches upon the expectations related to the Digital Services Act[1], (hereinafter referred to as the DSA), Artificial Intelligence Act[2] (hereinafter referred to as the AI Act), and the Directive 2024/1385/EU on combating domestic violence and violence against women adopted at the European Union (EU) level and how these instruments could advance a more inclusive and feminist approach to creating digital products, services, and policies.

Thus, we explore the possibility of a broad framework of indicators or benchmarks to monitor the integration of a feminist perspective in digital policies, specifically in terms of preventing cyber violence and online hate speech, improving content moderation, and empowering vulnerable individuals and groups. The framework could help ensure better reporting by tech companies in compliance with their existing obligations and support evidence-based and data-informed policy- and decision-making. Hence, a sub-objective of the study is to identify fields and topics, emerging from problems and gaps recognized by civil society and advocacy experts, where further research, practical efforts, legislation, and policies are needed.

If we adopt an optimistic perspective, 2024 can be seen as a pivotal year in Europe following the adoption or entry into force of several long-awaited documents aimed at, broadly speaking, protecting the rights of European Union (EU) citizens in an increasingly digital world, run by and through technology. The EU Digital Services Act (DSA) entered into force in February 2024, followed by the EU AI Act in August 2024 – most of the latter's rules to be applicable in

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

[2] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

two years' time, i.e. in August 2026. At the Council of Europe level (CoE), the Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law, the first legally binding convention in this field, was opened for signature in September 2024. Coming back to the EU, the Directive on combating violence against women and domestic violence, seemingly not directly related to the digital world, was adopted in May 2024. However, some of its key provisions oblige the member states to criminalize non-consensual sharing of intimate or manipulated material (Art. 5), cyber stalking (Art. 6), cyber harassment (Art. 7), cyber incitement to violence or hatred (Art. 8).

However, there are also those who approach these developments with less enthusiasm, pointing out the vagueness of some provisions or the difficulty of ensuring maximum enforceability or having all member states and the tech companies fully compliant. In the next section, we will return to and discuss in greater detail the key promises of these documents, as well as the main concerns of the more skeptical observers.

In this paper, we seek to address two research questions – or, rather, research puzzles. Firstly, we seek to understand what a feminist approach to the design and implementation of digital policy may entail, especially in view of the obligations to member states and companies set by the above-mentioned documents. Secondly, we attempt to identify the ways in which utilizing a feminist lens to digital policy may increase the enforceability of existing obligations, improve the monitoring of these documents' implementation, and contribute to decreasing the number, scope, and impact of potential harms in the digital world. To narrow down the topic of this study, we will concentrate on provisions and required steps needed to curb CVAW and online hate speech.

These questions were inspired and informed by ongoing calls for a feminist digital policy, which highlight its concern for and focus on human rights and privacy protection, the need to eliminate bias and intersectional discrimination and to engage marginalized groups. Thus, we must begin with several conceptual and theoretical clarifications.

2. Conceptual Framework

Importantly, there appears to be a consensus that digital policy is «not a stand-alone policy area, but lies across established political sectors such as health, education or the economy» (GLOTZBACH n.d.). It is commonly described as aiming to utilize the benefits and resources of digitalization while simultaneously tackling the challenges arising out of the use of digital technologies, including privacy and cyber risks or their potential negative impact on the environment (GLOTZBACH n.d.). When discussing digital policy, scholars and experts usually discern several key associated areas, issues, targets, or pillars. For instance, the EU policy program "Path to the Digital Future" sets specific targets in four areas: strengthening digital skills and education, secure and sustainable digital infrastructures, digital transformation of businesses, and digitalisation of public services (COUNCIL OF THE EU 2022). The policy was accompanied by a European Declaration on Digital Rights and Principles, which laid out the following principles to guide digital transformation in the EU: placing people and their rights at the center, solidarity and inclusion, freedom of choice, participation, safety and security, and sustainability (EUROPEAN COMMISSION 2024a). Offering a broader perspective, to the four thematic areas of digital policy proposed by Braman – access to the internet, access to content, copyright, data protection and the digital public sphere – Schmidt adds Ganz's suggestion of a fifth one, namely the digital public sphere (SCHMIDT 2021, 1).

Before presenting Schmidt's notion about what feminist digital policy should be characterized with, we should briefly mention some of the key values, positions, and premises linked to feminism, which are also applicable to policymaking and policy implementation. Kinsella contends that feminism's primary concern lies with «equality, justice, and the elimination of women's subordination and oppression» (KINSELLA 2023, 152) and that the birth of feminist theories of

international relations has «prompted a critical analysis of the existing discipline» (KINSELLA 2023, 54). As main trends associated with a feminist approach to foreign policy, Thomson identifies «a focus on sexual and reproductive health and rights (SRHR) and women's leadership; a clear commitment to existing international and human rights treaties and agencies; and, especially from civil society, an urge for policymakers to address the intersectional nature of gender (in)equality» (THOMSON 2022, 175). The centrality of gender, equality, and intersectionality is asserted also by Scheyer and Kumskova, who remind that feminist institutions and policies «function in accordance with feminist values, such as power-sharing, non-hierarchy and intersectional inclusiveness», which conditions feminist governance as based on gender equality, the application of an intersectional lens to policies, the reference to diverse gender identities, and the quest for transformation of traditional governance (SCHEYER, KUMSKOVA 2023, 239). Studying global trade governance, Hannah, Roberts and Trommer isolate four key feminist values that should inform it, namely «attentiveness to (1) structural inequality; (2) impacts of trade on different groups of people in multiple roles; (3) benefits for the social reproduction of people and communities; and (4) inclusivity and democracy» (HANNAH, ROBERTS, TROMMER 2023, 251).

Pinpointing if and how such values could be applicable to digital policy in various sectors and fields is not only possible and advisable, but it is also underway. Referring to how digital violence and hate speech disproportionately affect certain groups, Ehmke argues that a central question being asked by feminist digital politics with respect to a given service is who benefits from it, who is excluded from it and who is harmed by it (ROBERT BOSCH STIFTUNG 2022). Lindinger and Kloiber rightly note that a feminist digital policy is concerned with the effect of digital technologies on social inequalities – for instance, AI clearly has the potential to discriminate (directly or through proxy) based on gender, race, or disability, whilst the digital infrastructures' creation and sustainment often rely on exploitative labor in Global South extractive economies (LINDINGER, KLOIBER 2023). For Schmidt as well, digital policy needs to adopt a feminist, intersectional perspective to uncover «how forms of discrimination based on gender, social background or *race* are interconnected with new technologies and digital cultures» (2021, 2). Furthermore, she stresses that to be effective, feminist digital policy should «address the structures through which dominance is perpetuated and any use of the internet is influenced» and «focus on (ongoing) structures of discrimination and dominance» (LINDINGER, KLOIBER 2023, 7).

Therefore, drawing from the growing literature on the topic, at the center of a broader understanding of feminist digital policy we will find the pursuit of inclusivity and gender equality, the concern for existing bias and (intersectional) discrimination and the realization that meaningful change is not possible without dismantling larger oppressive and discriminatory structures.

3. *Methodology*

In this paper, we explore what a feminist approach to digital policy at the EU level may entail and whether it may have already been adopted, at least to some extent, in existing legislation or policy documents or by concrete actors with decision- and policymaking power. Hence, we perform in-depth systematic review and qualitative document analysis of existing legislation and policy and strategic documents to examine how and if they firstly address cyber violence against women and hate speech and secondly, embrace some of the topics and values identified as central to a feminist perspective on digital policy. The documents we study at the DSA, the EU AI Act, the EU Directive on combatting violence against women and domestic violence and the CoE Framework Convention[3]. Next, we examine and discuss established policies and relevant reports produced by

[3] Hereafter, for brevity, referred to DSA, AI Act, Directive on VAW and DV, and CoE Framework Convention.

several major social media companies and the way they approach violence against women and hate speech. For instance, we explore whether they acknowledge these phenomena as harmful, as well as the instruments and procedures related to content moderation, reporting, penalizing culpable users, etc., which these actors employ. We also refer to recent and relevant analyses, guidelines, and recommendations from experts and practitioners regarding how shortcomings in existing legislation and policies can be overcome and what improvements can be made, especially from a broadly feminist point of view and its concern for achieving inclusivity, non-discrimination, and intersectionality.

4. *Discussion*

4.1 *Review of Legislation, Policy, and Strategy*

The choice to focus on CVAW and online hate speech is motivated by the fact they are, firstly, wide-encompassing and cover different types of behavior, and secondly, and unfortunately, they are quite prevalent in digital spaces (probably, this is especially true about hate speech). Therefore, we start with a brief overview of the four landmark documents mentioned in the previous chapter to explore if and how they address specifically the two phenomena we concentrate on. Based on their contents, focus, and purpose, it is understandable that the Directive on VAW and DV and the CoE Framework Convention offer many more references to women's rights, equality and systemic inequalities, human rights and human dignity, etc. Importantly, all four documents contain statements, provisions, and remarks about the increased risk of discrimination in the digital world and the AI Act even warns against systems that may perpetuate «historical patterns of discrimination» including against women (EUROPEAN PARLIAMENT AND COUNCIL 2024a). On numerous occasions, the Directive on VAW and DV highlights the importance of intersectionality, as victims of intersectional discrimination are often at a heightened risk of violence, including cyber violence and hate speech.

Although the four documents are very recent, researchers, policy analysts, legal experts, and practitioners are already commenting on some imperfections of the provisions and discussing their implementability and enforceability. Looking at some of these insights could assist us in defining how the feasibility of certain obligations could be increased if we adopt a feminist approach.

As we realize it is not feasible to offer an in-depth presentation of the above-mentioned documents, we present those provisions and postulates which are most relevant to CVAW and online hate speech[4].

Digital Services Act (DSA)

Together with the Digital Market Act, it aims to «create a safer digital space in which the fundamental rights of all users of digital services are protected» (EUROPEAN COMMISSION 2025) and has specific rules and the strictest obligations for very large online platforms and search engines (VLOPSE[5]). With respect to CVAW, the document mentions that providers of VLOPSE should «diligently meet all their obligations… in respect of illegal content constituting cyber violence, including illegal pornographic content…» (EUROPEAN PARLIAMENT AND COUNCIL 2022, Recital 87). One of the risks mitigating measures specifically refers to cyber violence, while gender-based violence is mentioned with respect to the fourth category of risks

[4] Below we present excerpts from the mentioned documents, links to which are included in the Bibliography section.

[5] Those that have >45 million users per month in the EU.

and the contents of risk assessments to be performed by VLOPSE (*ibid*., Recital 83). With respect to hate speech, DSA states that the first systemic risk to be assessed by VLOPSE «concerns the risks associated with the dissemination of illegal content, such as the dissemination of child sexual abuse material or illegal hate speech…» (*ibid*., Recital 80). Furthermore, one of the measures to mitigate this risk is «adapting content moderation processes…and, where appropriate, the expeditious removal of, or the disabling of access to, the content notified, in particular in respect of illegal hate speech or cyber violence…» (*ibid*., Article 35 (c)).

EU AI Act

This Act introduces new rules on artificial intelligence and aims «to foster trustworthy AI in Europe and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models» (European Commission, no date). The document does not explicitly mention cyber violence against women, harassment, abuse, or hate speech. However, it requires that certain AI systems to be deployed in education and employment should be classified as high-risk, because they may «violate the right not to be discriminated against and perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation» (EUROPEAN PARLIAMENT AND COUNCIL 2024a, Recitals 56 and 57).

EU Directive on Combating VAW and DV

The Directive recognizes that «(e)quality between women and men and non-discrimination are core values of the Union and fundamental rights enshrined, respectively, in Article 2 of the Treaty on European Union (TEU) and in Articles 21 and 23 of the Charter of Fundamental Rights of the European Union (the 'Charter')» and, as such, are endangered by VAW and DV (EUROPEAN PARLIAMENT AND COUNCIL 2024b, Recital 2). With respect to CVAW, the document explores in depth phenomena such as violence against women, cyber violence, domestic violence as well as their implications for gender equality, women's rights, etc. Importantly, the Directive calls on states to criminalize female genital mutilation, forced marriage, the non-consensual sharing of intimate or manipulated material, cyber stalking, cyber harassment, cyber flashing and cyber incitement to violence or hatred. Regarding hate speech, the Directive asserts that in the digital world, hate speech «is reinforced by the online disinhibition effect, as presumed anonymity on the internet and a sense of impunity reduce people's inhibition to engage in such speech» and that «Women are often the target of sexist and misogynous hate online, which can escalate into hate crime offline» (*ibid*., Recital 25).

CoE Framework Convention

According to the Convention's provisions, AI systems must comply with human dignity and individual autonomy, equality and non-discrimination, respect for privacy and personal data protection, transparency and oversight, accountability and responsibility, reliability and safe innovation. State Parties should «adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and potential impacts to human rights, democracy and the rule of law» (COUNCIL OF EUROPE 2024, Article 16(1)). The Convention does not explicitly mention cyber violence against women and hate speech but, as mentioned above, lays out in detail the commitments ensuing from the document that are related to certain fundamental principles. It should be noted that the Convention is a much shorter document than the ones produced within the EU framework.

When exploring whether the impact assessments envisioned in the DSA could improve protection from technology-facilitated gender-based violence (GBV), Martins acknowledges that DSA treats GBV as a systemic risk but argues that referring to it in a way unconventional to international law on human rights and gender shifts the focus from rights assessment to risk assessment, which may erode efforts to tackle GBV and other rights violations (Martins 2024, 8). The author also echoes concerns about human rights impact assessments (HRIAs) in general, namely that they often do not include a gender dimension and employ gender stereotypes against women and fail to consider gendered power dynamics (Martins 2024, 10). Very concrete recommendations pertaining to gender and GBV were formulated by the European Institute for Gender Equality (EIGE) during a public consultation on the design of transparency reports under the DSA. EIGE emphasizes «the need to include gender-based measurement instruments», which, for example, would mean recognizing the prevalence of gender-based hate speech and the «introduction of separate categories to measure bullying, harassment, and stalking on the grounds of gender» (EIGE 2024). A September 2024 report voices similar concerns about DSA shared by others – that the document does not recognize the "gendered nature of harm", that it does not clearly define what is harmful or illegal (leaving it to the member states to use their own definitions) and that the lack of a common understanding what illegal is «renders the DSA almost unfit for purpose in the face of digital and online violence» (EWL 2024, 59).

The AI Act has been subjected to similar criticisms. According to De Vido, the document contains «no reference to the disproportionate impact of AI-generated gender-based violence on women and girls, except for a very short reference in the preamble...» and despite introducing minimum standards for foundational models, «omits any content moderation» (De Vido 2024, 4). Elsewhere, it has been suggested that AI and gender should be further interconnected through and in legislation – while the EU has adopted both the AI Act and the new Directive on VAW and DV, these «should also be combined with legislation aimed specifically at AI-created abuses» (de Alwis, Vialle 2024). Lütz argues that, given the focus on human rights protection, the AI Act «could have addressed more concretely some of the underlying gender and non-discrimination issues» and that these shortcomings «will have to be addressed via new instruments specifically designed for gender equality and non-discrimination or by revising existing instruments in the light of new technological developments» (Lütz 2024).

The adoption of the EU Directive on VAW and DV has been praised by many experts in countering VAW, policymakers, practitioners, and civil society organizations. The criminalization of certain behaviors in digital spaces is indeed an important step, which, paired with stepped-up prevention efforts and greater collaboration among different stakeholders, has the potential to reduce cyber VAW, deter perpetrators, and ensure better protection of victims. However, a number of valid criticisms have been put forward along with calls to revise and perfect the document in several years' time. Two identified shortcomings are that the articles on cyber VAW currently place the burden of proof on the victim as they refer only to intentional conduct and that new forms of violence perpetrated by AI are not specifically mentioned in the said articles (EWL 2024, 52, 54). In addition, Allen and Thakur assert that while the Directive has called for the criminalization of serious forms of cyber VAW, it did not cover other forms «including actions that may not rise to the level of illegal conduct, but which can nevertheless have a chilling effect on women and non-binary people's speech» (Allen, Thakur 2024). Now, when the Directive and the DSA have entered into force, it is also crucial to «ensure that human-rights centred approach to enforcement and high standards of accountability towards due diligence obligations are maintained» (Allen, Thakur 2024).

Digital policy experts have expressed concerns that the CoE Framework Convention offers a reaffirmation of existing practices instead of new substantive regulatory measures, which may undermine its effectiveness. It has been also highlighted that allowing exemptions for private actors, not only in the context of the Convention but also with respect to any AI regulation,

increases the risk of human rights violations, especially considering the heightened vulnerability of certain groups such as women, non-binary, and LGBTQ+ people, and racialized communities (LEUFER 2024).

4.2 *Frameworks and Solutions by Social Media Companies*

Usually, when exposure to online hate speech, abuse, and bullying is discussed, it is the social media companies/tech giants that are blamed for not doing more to protect their users' safety, privacy, and well-being. While most of the content on social media is created by users (meaning human beings) and not by platforms, , companies could, and some say should, correct harmful or illegal behavior by improving content moderation, blocking the monetization of such content, and being more accountable and transparent with respect to how their services and products are used. They should also comply with their obligations existing under current regulatory frameworks, legislation, etc. and are encouraged to go beyond these obligations to demonstrate in practice their commitment to fundamental rights and freedoms, etc. In the United States, ongoing discussions surrounding section 230 of the 1996 US Communications Decency Act, a key provision, which relates to companies' immunity from liability for content posted by users[6], demonstrate the difficulty of whether and how service providers should be liable for hosting harmful content produced by others. Danielle Citrone recognizes that § 230 has encouraged expression by protecting platforms from liability for content generated by their users and allowed for the creation of numerous online services (CITRONE 2022, 717); however, it has also provided «a legal shield» for online abuse, stalking, harassment, privacy violations with most of the victims being women and people from marginalized groups (CITRONE 2022, 717-718). This is why Citrone is proposing to reform the text to prevent websites that «deliberately encourage, solicit, or maintain intimate privacy violations, cyber stalking, or cyber harassment» and make companies demonstrate they took actions to address abuse (CITRONE 2022, 713-714). Reviewing Citrone's proposal, Kosseff remains unconvinced that § 230 needs to be modified, as a change may cause more harm than good and reminds of the important political aspect – «the people who run Washington are deeply divided as to what they want the internet to look like and the role that they envision platforms playing» (KOSSEFF 2023, 774). Reflecting on the optimal solutions concerning liability for intermediaries and enforcement, Husovec states that «law-makers should not be technology prophets, but rather clever incentive-setters» as 100% enforcement could cause harm to society (HUSOVEC 2017, 222). Elsewhere, examining in detail the DSA, Husovec makes another key point – that regulation in liberal democracy should also be about empowering people and putting emphasis on individuals and civil society instead of resulting in a state-dominated-model replacing the private-power-dominated model (HUSOVEC 2024, 11). Crafting legislation and policies with this goal in mind will be consistent with the understanding that technology should be accountable and responsible, its creators mindful of the human rights framework.

In this sub-section, we will explore, by utilizing a gender lens, how two social media companies – Meta's Facebook and TikTok – have shaped, formulated, and monitored their policies and practices[7], as well as how they are reporting on their DSA obligations. Facebook remains the largest social media platform in terms of the number of active users (more than 3 billion in 2024), while TikTok is very popular among teenagers and younger people (technically, the minimum age for using TikTok is 13).

[6] For a detailed exploration of how Section 230 came to be, see Jeff Kosseff, The Twenty-Six Words that Created the Internet, Cornell University Press, Ithaca-London, 2019

[7] Links to the reviewed policies and reports by Meta and TikTok are provided in the Bibliography section below. This article examines the two companies' policies relevant as of November 2024.

Facebook/Meta

We reviewed the following documents pertaining to Facebook: the platform's Community Standards, more specifically the sections on Objectionable Content, Safety and Misinformation, the latest of the quarterly published Community Standards enforcement reports (as of November 2024. this is the Q2 2024 report) and the three transparency reports prepared under the DSA obligations (covering the period April 2023 – September 2024). Drawing on the tenets of what a feminist digital policy may constitute, we identify several areas where meaningful and real improvements could increase the company's compliance with regulations and human rights standards.

One weakness of the Community Standards is that there is no special section, or a sub-section, dedicated to gender-based violence or violence against women. This is problematic considering that available data, albeit insufficient, shows women and girls are disproportionately affected by cyber violence. Specific provisions related to abuse on the platform targeting individuals at risk of intersectional discrimination are also needed to demonstrate awareness on the part of the company that these people are more vulnerable than others to certain harmful behaviors. The explored sections do not elaborate on the gendered nature of many conducts, which are not permissible on Facebook. For instance, the sub-section on Suicide, Self-Injury, and Eating Disorders could be expanded to acknowledge that girls and young women are at high risk of developing an eating disorder as they are exposed to unhealthy and unrealistic beaty standards on social media. The part of Misinformation could also be enhanced by recognizing that women and LGBTIQ+ people, especially those in visible positions are increasingly vulnerable to online misinformation and disinformation, which is often intended to mock their work, silence their voices, cause distress or force them out of the social platform altogether. Overall, Meta's choice to talk about misinformation and not disinformation is intriguing and raises the question of why in this segment of the Community Standards, the company does not refer to the more organized and intentional spreading of false or misleading information. Removing misinformation that may cause physical harm is a priority and as much as this is understandable, considering that tremendous amount of content reaches billions in seconds, acknowledging that misinformation could also cause emotional and mental suffering to people is crucial and indispensable. It is noteworthy that the section on Hate Speech provides a detailed list of what Meta defines as Protected Characteristics – race, ethnicity, sex, etc. – based on which people may become attacked online. However, while the list includes 'gender identity', it does not mention 'gender' alone, which falls short of stating explicitly that some Facebook users are at risk of hate speech because of stereotypical socially constructed notions about a woman's and a man's place and role in society, which are often transferred to the virtual world.

The goal of Meta's Community Standards enforcement report is «to more effectively track» the company's progress and demonstrate its «continued commitment to making Facebook and Instagram safe and inclusive» (META 2024). As explained by Meta, the collected data underlying these reports shows the prevalence of, for example, hate speech, the number of pieces the company acted on or restored, etc. However, these datasets do not shed light on (at least) the particular 'protected characteristic' based on which a piece of content is considered hate speech or whether the actioned content was in a language different than English, was it textual or visual, etc. For researchers and policymakers would be valuable to explore on how many occasions hate speech was directed towards or reported by users who identify themselves as women or non-binary – obviously, without revealing any sensitive data about the users.

Since the publication by VLOPSEs of the first transparency reports required under the DSA, experts and observers have expressed legitimate concern that tracking, assessment, and comparison are made difficult by the «lack of transparency reporting standardization, varying degrees of disclosure, and seemingly different interpretations of DSA Articles that VLOPs and

VLOSEs are obligated to comply with» (MILLER 2023). Similar criticism led the European Commission to commence work on a standardized reporting template and a public consultation on its draft took place in the beginning of 2024. Consequently, in the beginning of November 2024, the Commission adopted an implementing regulation with a quantitative and a qualitative templates transparency reporting and a harmonized reporting periods – an action that could be considered a positive development. It is commendable that an entire category of the template includes labels for Cyber Violence against Women covering, among others, cyber harassment against women and gendered disinformation. Furthermore, the instructions for filling in the templates specify that a specific category like 'Illegal incitement to violence and hatred against women' should always take precedence over the more generic one 'Illegal incitement to violence and hatred based on protected characteristics (hate speech)' (EUROPEAN COMMISSION 2024b). VLOPSEs could also be proactive and emphasize their commitment to human rights and responsible technology by specifying, in the qualitative section of the template, the steps taken to eliminate algorithmic bias in automated tools used for content moderation – particularly with respect to gender, race, and other 'protected characteristics'. In addition, the section on human resources dedicated to content moderation may contain data about the percentage of women in this workforce. Reports could also benefit from more detailed information about the psychological support provided to human content moderators.

Lastly, it should be noted that Meta published annual human rights reports, which examine the company's progress towards its commitments under the UN Guiding Principles and Human Rights and its own Corporate Human Rights Policy. These documents do mention the company's commitment to addressing gender-based harassment, gender-based violence, and attacks against women running in elections, journalists, and human rights defenders. However, more high-quality and reliable data is needed to study and assess how these and similar pledges are carried out in practice and whether they have the intended and desired effect.

TikTok/ ByteDance

In the TikTok's Safety Center, we can find topics such as Countering Hate Speech & Behavior, Sexual Abuse Support, Bullying Prevention among others; however, it deserves noting that this initial list of topics related to safety is much shorter than the one in Facebook's Community Standards. In contrast to Facebook, this platform's description of 'protected attributes' includes both gender and gender identity. Similarly to Facebook, TikTok fails to recognize that certain groups may be more vulnerable to online risks due to a number of factors, frequently because of their intersecting identities. The Safety Center does not identify women and girls as a group that deserves special protection as it is disproportionately targeted by harmful online behavior – cyber harassment, hate speech, and stalking. In general, the descriptions of the topics could be expanded to elaborate on how some phenomena line cyber violence or body shaming can be gendered or how violence against women is often rooted in historical and systemic inequalities that place women and girls in subordinate positions. Among the ten main topics in the Safety Center, we also do not find disinformation or/and misinformation – although the company has published statements and reports on its counter-efforts, the challenge these phenomena present to women, especially those in public-facing positions should be acknowledged more unambiguously. The concern for their users' well-being – and this applies to both Facebook and TikTok – could be highlighted by providing information (definitions and possibly examples) about the different *types* of cyber abuse that women, girls, and vulnerable groups may experience (doxing, trolling, non-consensual sharing of image, etc.). For that purpose, the companies may refer to existing definitions offered by the European Union, Council of Europe, UN Women, the European Institute for Gender Equality, etc.

Two sections of the TikTok's Community Guidelines Enforcement Reports are noteworthy,

the first one being the segment, which shows the removal volumes and rates, by market and the human moderation language distribution. The second segment shows the distribution of removed videos by the platform policy that was violated – this allows us to see that in the period April – June 2024, the biggest share of content removals (31%) was related to Sensitive and Mature Themes (of which 23% Sexually Suggestive Content, 22.3% Nudity and Body Exposure, 21% Sexual Activity and Services, 19.6% Shocking and Graphic Content, and 14% Animal Abuse). Again, both the social platform and researchers studying harmful online behavior, could benefit from more comprehensive and high-quality data about the removed content. It is also crucial, when analyzing compliance with the community guidelines or drafting policies to increase inclusivity and improve well-being, to consider that many negative phenomena are gendered in nature and stem from deeply engrained notions that certain sexes, genders, ethnicities, etc. are inferior to others.

Three reports on TikTok's compliance with the DSA have been prepared and published as required by the act. Their format and contents are slightly different from the ones submitted by Meta, which again underscores the importance of utilizing a standardized template. Hopefully, this will allow both researchers and experts and the social platforms' users to have a better understanding about the prevalence of harmful and illegal behavior, the main groups targeted by it and, most importantly, about the adopted countermeasures.

At the end, as of November 2024, the European Commission has already opened formal proceedings (two against TikTok and one against Meta) under the DSA. The first proceedings against TikTok aim to investigate whether the social platform provides adequate protection to minors, data access to researchers and risk management of addictive design and harmful content. The basis for the second proceedings is TikTok's failure to submit a risk assessment report before launching TikTok Life in France and Spain. The proceedings against Meta (Facebook and Instagram) will examine the company's policies and practices relating to deceptive advertising and political content. The outcome of the investigations and the subsequent steps by the two companies to resolve breaches of obligations and ensure compliance with the DSA is crucial for ensuring their commitment to transparency, accountability, and the safety and well-being of their users.

5. *Conclusion – Future Research, Action, and Advocacy*

This paper explored the provisions of existing legislation, applicable to the efforts to counter CVAW and online hate speech, as well as the current policies and reported actions of several media platforms again with a focus on these two malicious phenomena. Our main goal was to argue that adopting a feminist perspective to the drafting, review, and implementation of these and similar documents could strengthen their effectiveness when it comes to safeguarding the rights of women, children, and vulnerable communities and individuals in general. Considering that gender inequalities and other systems of oppression still pervade societies; it is essential to reflect on how digital policies could embrace feminist principles and values.

Several ways to do that could be deducted from the critiques and recommendations made by the experts studying both the documents produced at the EU and CoE level and the guidelines, policies, and reports published by social media platforms. Among these stand out the need to prioritize rights assessments over risk assessments and the importance of recognizing the existence of specific and intersecting vulnerabilities. Furthermore, it is essential for policymakers, but especially for companies, to accept and act upon the fact that cyber violence and hate speech can be gendered and, in fact, often are. For instance, hate speech targeting a woman or a group of public-facing women may be an expression of harmful gender stereotypes or misogynistic notions about women and their role in society. Similarly, cyber violence may

involve gender-based insults and the sexualization and objectification of women. It is also crucial to pursue a human rights-centered and victim-centered approach when formulating and executing policies and strategies, which also includes measures to avoid secondary victimization. Women, and victims of cyber violence and hate speech in general, should not be reprimanded for not using a platform's reporting function or not knowing how to use it 'properly'. In addition, as already mentioned before, the burden of proof should not fall on the victim of CVAW or hate speech, not least because it is likely to cause the victim to re-experience the abuse and, hence, re-live the trauma. The impetus to introduce and implement these and similar measures could be driven by the civil society sector, as good, clear, and meaningful advocacy is key.

Moreover, conceptualizing and introducing a comprehensive framework of indicators or benchmarks to monitor the integration of feminist-inspired and -informed principles in digital policies may be worth considering. Such a framework, stressing the prevention of CVAW and online hate speech, the improvement of content moderation, and the empowerment of vulnerable individuals and groups could be initiated at a regional (EU, CoE, or OSCE) or global (UN) level. The framework could help ensure better reporting by tech companies in compliance with their existing obligations and support evidence-based and data-informed policy- and decision-making. Such a framework could be built based on the active engagement of the main stakeholders, namely those most at risk of experiencing CVAW and online hate speech. Hence, the process of its creation could involve intersectional analysis, gender analysis, extensive consultations with stakeholders (including victims and marginalized groups and tech experts), qualitative and quantitative analysis of recent, reliable and comprehensive data about the current prevalence and features of CVAW and online hate speech, as well as the creation of a thorough methodology to monitor and evaluate the application of these indicators/benchmarks. The introduction of a harmonized template for the DSA transparency reports by the European Commission could be viewed as a promising step in this direction. However, its effectiveness is conditional upon the good monitoring of its utilization.

Lastly, those applying a feminist lens to digital policy can explore the lessons learnt (achievements and shortfalls) from the application of feminist principles in other fields, such as foreign policy and international development. It is essential to explore the potential of a feminist digital policy to eliminate inequalities and bias-perpetuating structures and uphold human dignity and non-discrimination, while countering malicious phenomena in the cyber world.

Bibliography

ALLEN A., THAKUR D. 2024. *CDT Europe Reacts to EU Directive on Gender-Based Violence (GBV) – New Rules to Tackle Online GBV Create Free Expression Concerns*, Center for Democracy&Technology. Available at: https://cdt.org/insights/cdt-europe-reacts-to-eu-directive-on-gender-based-violence-gbv-new-rules-to-tackle-online-gbv-create-free-expression-concerns/ (accessed 4/11/2024).

CITRON D.K. 2022. *How To Fix Section 230*, in «Boston University Law Review», Virginia Public Law and Legal Theory Research Paper No. 2022-18. Available at SSRN: https://ssrn.com/abstract=4054906 (accessed 2/3/2025).

COUNCIL OF EUROPE 2024. *The Framework Convention on Artificial Intelligence*. Available at: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence (accessed 25/11/2024).

COUNCIL OF THE EU 2022. *'Path to the Digital Decade': Council adopts key policy programme for EU's digital transformation*. Available at: https://www.consilium.europa.eu/en/press/ press-releases/2022/12/08/path-to-the-digital-decade-council-adopts-key-policy-programme-for-eu-s-digital-transformation/ (accessed 27/10/2024).

DE ALWIS DE SILVA R., VIALLE E. 2024. *Is AI-Facilitated Gender-Based Violence the Next Pandemic?* The Regulatory Review. Available at: https://www.theregreview.org/2024/05/06/de-silva-de-alwis-vialle-is-ai-facilitated-gender-based-violence-the-next-pandemic/ (accessed 16/11/2024).

DIGWATCH 2024. *Council of Europe opens AI convention for signature*. Available at: https://dig.watch/updates/council-of-europe-opens-ai-convention-for-signature (accessed 3/11/2024).

DE VIDO S. 2024. *Deep Fake as AI-generated Violence against Women*. DEP Deportate Esuli Profughe. Available at: https://www.unive.it/pag/fileadmin/user_upload/dipartimenti/DSLCC/documenti/DEP/finestra/Finestra_19.pdf (accessed 16/11/2024).

EUROPEAN COMMISSION (n.d.). *Strengthening online platforms' responsibility*. Available at: https://commission.europa.eu/topics/countering-information-manipulation/ strengthening-online-platforms-responsibility_en#:~:text=The%20Artificial%20Intelligence%20(AI)%20 Act,-The%20 AI%20 Act&text=The%20aim%20of%20the%20new,powerful%20and %20 impactful%20AI%20models (accessed 30/3/20205).

EUROPEAN COMMISSION 2025. *The Digital Services Act package*. Available at: https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (accessed 30/3/2025).

EUROPEAN COMMISSION 2024a. *European Digital Rights and Principles*. Available at: https://digital-strategy.ec.europa.eu/en/policies/digital-principles#tab_1 (accessed 7/11/2024)

EUROPEAN COMMISSION 2024b. *Digital Services Act – transparency reports (detailed rules and templates)*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14027-Digital-Services-Act-transparency-reports-detailed-rules-and-templates-_en (accessed 26/11/2024).

EUROPEAN INSTITUTE FOR GENDER EQUALITY (EIGE) 2024. *EIGE stresses the need for gender dimension in DSA transparency reporting*. Available at: https://eige.europa.eu/ newsroom/news/eige-stresses-need-gender-dimension-dsa-transparency-reporting?language_content_entity=en (accessed 16/11/2024).

EUROPEAN PARLIAMENT AND COUNCIL 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). Available at: https://eur-lex.europa.eu/eli/reg/2022/2065/oj (accessed 25/11/ 2024).

EUROPEAN PARLIAMENT AND COUNCIL 2024a. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689 (accessed 25/11/2024).

EUROPEAN PARLIAMENT AND COUNCIL 2024b. Directive 2024/1385 Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence. Available at: https://eur-lex.europa.eu/legal-content/EN-BG/TXT/?from=EN&uri=CELEX%3A32024 L1385 (accessed 27/10/2024).

EUROPEAN WOMEN'S LOBBY (EWL) 2024. *Report on Cyber Violence against Women.* Available at: https://www.womenlobby.org/IMG/pdf/ewl_fullreport_cvawg.pdf (accessed 16/11/2024).

GLOTZBACH S. (n.d.). *Digital Policy*. Available at: https://www.giz.de/expertise/html/60114.html#:~:text=Digital%20policy%20is%20not%20a,health%2C%20education%20or%20the%20economy (accessed 24/6/2024).

HANNAH V., ROBERTS A., TROMMER S. 2023. *Feminist interventions in trade governance*, in SAWER M., BANASZAK L.A, TRUE J., KANTOLA J. (eds.), *Handbook of Feminist Governance,* Edward Elgar Publishing, 250 ff.

HUSOVEC M. 2024. *Principles of the Digital Services Act,* Oxford University Press.

HUSOVEC M. 2017. *Cambridge Intellectual Property and Information Law: Injunctions Against Intermediaries in the European Union Accountable but not liable?*, Cambridge University Press.

KINSELLA H.M. 2023. *Feminism,* in BAYLIS, J., SMITH, S., OWENS, P. (eds.), *The Globalization of World Politics an Introduction to International relations*, Oxford University Press, 9th ed., 147 ff.

KOSSEFF, J. 2023. *What Was the Purpose of Section 230? That's a Tough Question, a Response to Danielle Citron's How to Fix Section 230,* in «Boston University Law Review». Available at SSRN: https://ssrn.com/abstract=4388216 or http://dx.doi.org/ 10.2139/ ssrn.4388216.

LEUFER D. 2024. *Why human rights must be at the core of AI governance*, AccessNow. Available at: https://www.accessnow.org/human-rights-and-ai-governance/ (accessed 4/11/2024).

LINDINGER E., KLOIBER K. (2023). *Exploring intersections: a feminist perspective on digital and foreign policy. Heinrich Boll Stiftung*. Available at: https://eu.boell.org/en/2023/06/01/ feminist-digital-foreign-policy (accessed 17/6/2024).

LÜTZ F. 2024. *The AI Act, gender equality and non-discrimination: what role for the AI office?*, in «ERA Forum», 25, 2024, 79 ff. Retrieved Nov 17, 2024 from https://doi.org/10.1007/s12027-024-00785-w.

MARTINS P. 2024. *How Can Impact Assessments Improve Protection from TFGBV?* Centre for International Governance Innovation. Available at: https://www.cigionline.org/static/documents/DPH-paper-Martins.pdf (accessed 20/6/2024).

META (n.d.). *Community Standards.* Available at: https://transparency.meta.com/policies/community-standards/ (accessed 18/11/2024).

META 2024. *Community Standards Enforcement Reports.* Available at: https://transparency.meta.com/ reports/community-standards-enforcement/ (accessed 18/11/2024).

MILLER G. 2023. *First Transparency Reports Under Digital Services Act Are Difficult to Compare,* Tech Policy Press. Available at: https://www.techpolicy.press/first-transparency-reports-under-digital-services-act-are-difficult-to-compare/ (accessed 5/11/2024).

ROBERT BOSCH STIFTUNG 2022. *"Society as a whole will benefit from feminist digital politics".* Available at: https://www.bosch-stiftung.de/en/storys/society-whole-will-benefit-feminist-digital-politics (accessed 17/6/2024).

SCHEYER V., KUMSKOVA M. 2023. *Feminist peace and security governance and the UN Security Council*, in SAWER M., BANASZAK L.A, TRUE J., KANTOLA J. (eds.), *Handbook of Feminist Governance*, Edward Elgar Publishing, 238 ff.

SCHMIDT F. 2021. *Digital Policy: A feminist Introduction.* Available at: https://eu.boell.org/en/2021/02/15/digital-policy (accessed 24/6/2024).

TIKTOK 2024. *Community Guidelines Enforcement Report*. Available at: https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement-2024-9 (accessed 25/11/2024).

THOMSON J. 2022. *Gender norms, global hierarchies and the evolution of feminist foreign policy*, in «European Journal of Politics and Gender», 5(2), 173. Retrieved Nov 25, 2024, from https://doi.org/10.1332/25151082lX16354220233761.

Funding

The research for this paper has been carried out within the framework of the project Stereotyping, Disinformation, and Politicisation: links between attacks against the Istanbul Convention and increased online gender-based violence (RESIST), which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945361. This paper reflects only the author's view and the Agency and the Commission are not responsible for any use that may be made of the information it contains.