

(Legal) Issues Concerning the Content Moderation on Social Networks

Dr. Kristina ČUFAR

University of Ljubljana, Faculty of Law, Ljubljana (Slovenia)

(TFL Glasnik, 34/2021, ISSN 2350-5753)

Social networks are changing the world. They have radically altered the way their users connect and communicate with each other, as well as the way they access and exchange information. These transformations are a double-edged sword. On the one hand, social networks play an important role in connecting (formerly) socially isolated individuals, offer the possibility of obtaining direct information about otherwise overlooked events (e.g., human rights violations in conflict areas) and enable networking of activist groups organizing demonstrations, charity events and awareness-raising for a wide variety of societal issues. On the other hand, the romantic dream of the Internet as a boundless field of exchange of information, pluralism, horizontal integration and self-regulation by users that marked the 1990s,¹ is fading in the privatization and monopolization of the World Wide Web by corporations. The role of social networks in disseminating misinformation and disinformation, conspiracy theories, copyright infringements, echo chambers, hate speech, harassment, intimidation and extortion, and other problematic phenomena stimulates debates about the inadequacy of current regulation.

These negative phenomena are related to the algorithmic structure of social networks to some extent. Social network companies want to keep users on their platforms for as long as possible and show them targeted ads. Social networks mainly show users posts that might attract them and conform with their views and beliefs.² This poorly regulated aspect of social networks is often associated with increasing polarization in society and increasingly harsh and offensive rhetoric. In recent years, we are facing declining levels of appropriateness of expression of politicians and other public figures – as well as everyone else on social networks. Offensive expression seems to be replacing arguments and traditional media pay a lot of attention to such expressions of opinions and ideas on social networks, thus spreading it far beyond their platforms. Slovenian courts are increasingly confronted with cases involving problematic and

¹ John Perry Barlow: A Declaration of the Independence of Cyberspace, in: Electronic Frontier Foundation, 1996; <<https://www.eff.org/cyberspace-independence>> (30. 11. 2021).

² Aleš Rozehnal: The Role of Social Media in Shaping Society, in: M. Wielec: The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism. Ferenc Mádl Institute of Comparative Law, 2021, pp. 217–276.

potentially illegal forms of expression on social networks and occasionally even recognize it as a specific way of expression inherent in these networks.³ Of course, courts faced with the difficult task of balancing the rights of two parties (e.g., freedom of expression on the one hand and personal rights on the other) do not always take this path.⁴ The role and importance of social networks in judicial decision-making does not yet seem to be fully consolidated. However, the issue of inappropriate expression on social networks certainly goes beyond the bare issue of legally permissible expression, as its social consequences are wide and worrying.⁵ Nevertheless, issues related to social networks, especially when concerning freedom of expression and its abuse on social networks' platforms require thorough reflection and regulatory responses.

Companies that offer social networks decide on the content displayed to individual users and, moreover, also decide on the removal of inappropriate content in accordance with their own rules.⁶ Moderation of content on social networks is necessary, as a large number of users around the world upload huge amounts of problematic content to their platforms (e.g., videos of child and adult sexual abuse, images of violence, terrorist propaganda and extremism, glorification of eating disorders and suicide). However, the moderation of content provided by social media through artificial and human intelligence too often leads to controversial decisions (e.g. Facebook's removal of the Venus of Willendorf for violating the platform's naked nudity rules at the time).⁷ Such scandals highlight the issue of private control over publicly available content and the consequences it has for society as a whole.

Social networks remove not only content that represents (potential) violations of the legal provisions of individual countries, but also content that they determine as inadmissible. Defining specific content as inadmissible is a matter of self-regulation and sometimes involves

³E.g.: Judgement II Ips 75/2019, Supreme Court of the Republic of Slovenia, February 6, 2020. See also Ana Jereb: *Žaljive objave na družbenih omrežjih kot moderni izziv sodišč: so lahko tudi vsečki protipravni?*, v: *Pravni Letopis 2020*, 2020, pp. 175–199.

⁴Judgement II Ips 22/2021, Supreme Court of the Republic of Slovenia, September 1, 2021.

⁵This is not a novel problem, see e.g.: Mojca M. Plesničar in Pika Šarf: 'This Web Page Should Not Exist': A Case Study of Online Shaming in Slovenia, in: D. Trottier, Rashid Gabdulhakov in Qian Huang (eds.): *Introducing Vigilant Audiences*. Open Book Publishers, 2020, pp. 188–213.

⁶The practice of shadowbanning can be particularly problematic in this context: the user's posts are not deleted, but social networks do not show them to other users. András Koltay: *The Regulation of Social Media Platforms in Hungary*, in: M. Wielec (ed.): *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism*. Ferenc Mádl Institute of Comparative Law, 2021, pp. 92–93.

⁷Frederik Stjernfelt and Anne Mette Lauritzen: *Facebook's Handbook of Content Removal*, in: F. Stjernfelt in A. Mette Lauritzen (eds.): *Your Post Has Been Removed: Tech Giants and Freedom of Speech*. Springer International Publishing, 2020, pp. 115–137.

cooperation with countries and their organizations, in our context, the EU.⁸ This raises the question of whether leaving the decision on whether certain content is (potentially) illegal to social networks is the right way to address the problematic content on social networks. Legislative solutions, such as the German law known as the NetzDG, which require social networks to monitor and remove allegedly illegal content relatively briskly, are often criticized for encouraging preventive removal of content that isn't legally problematic, potentially leading to suppression of free debate.⁹ A slightly different attempt at state regulation of expression on online networks is proposed in Poland, which intends to impose state oversight over content moderation processes.¹⁰ The proposed legislation, which is supposed to address the problem of misinformation and violations of rights to social networks, provides for the establishment of a special body (Council for Freedom of Speech) to monitor the timely removal of misinformation and violations of freedom of expression by social networks. Such approaches are also not immune to criticism, but since the regulatory solution is still in a draft phase its effects and implementation cannot be analyzed yet.

In most EU Member States, including Slovenia, host service providers are not required to control and monitor content posted on their platforms, but only to remove or disable access to content as soon as they gain knowledge of its illegality.¹¹ Existing legislation is based on the e-Commerce Directive, which is over 20 years old and has to some extent lost touch with the development of digital technologies. For this reason, the European Commission proposed a Digital Services Act package in December 2020, which will address some of the shortcomings of the existing regime.¹² The final form of the proposed regulation is not yet clear, but its critics are already expressing their doubts.¹³ It is difficult to effectively regulate complex phenomena

⁸ European Commission: 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> (1. 3. 2021); European Commission: Code of Practice on Disinformation; <<https://ec.europa.eu/digital-single-market/en/code-practice-disinformation>> (1. 3. 2021).

⁹ Victor Claussen: Fighting Hate Speech and Fake News. The Network Enforcement Act (NetzDG) in Germany in the Context of European Legislation, in: *Media Laws*, nu. 3/2018.

¹⁰ Marcin Wielec: Censorship on Digital Platforms and Social Media Versus Freedom of Expression and Pluralism: The Perspective of the Republic of Poland, in: M. Wielec (ed.): *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism*. Ferenc Mádl Institute of Comparative Law, 2021, pp. 40–43.

¹¹ Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (e-Commerce Directive); *Zakon o elektronskem poslovanju na trgu* (ZEPT, Eng. 'Electronic Commerce Market Act') Official Gazette RS, No. 97/07.

¹² European Commission: The Digital Services Act Package, 2 June 2020; <<https://ec.europa.eu/digital-single-market/en/digital-services-act-package>> (30. 11. 2021).

¹³ Natali Helberger: The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power, in: *Digital Journalism* 8, nu. 6/2020, pp. 842–854.

related to social networks, such as dissemination of misinformation and disinformation, high incidence of hate speech and many other problematic content. The EU is certainly an actor that will determine the direction of the development of regulation and dealing with these phenomena in our part of the world. However, as the aforementioned examples of Germany and Poland show, individual Member States are paving their way in the field of content regulation on social networks.

While the legislative solutions put forward by the Western Member States are often perceived as a driving force in this area, approaches to regulating social networks in Central and Eastern European countries tend to receive less attention. A comparative legal excursion into the legal regulation of content on social networks certainly contributes to a better understanding of existing regulation and future trends in the EU and beyond. The Ferenc Mádl Institute of Comparative Law seeks to promote an in-depth study of the current state of legislation and its study in Central and Eastern European countries with an initiative to research the legal regulation of content moderation and misinformation and disinformation on social networks in the context of the Central European Professor's Network. Not necessarily limited to the EU Member States, it promotes dialogue between experts and provides a vivid insight into established legislative practices and the direction of their development. In the current climate in the EU, an insight into the scholarship of established lawyers from Eastern and Central Europe is certainly an important contribution to understanding the debates that may mark the coming years. Thanks to the Ferenc Mádl Institute's Central European Professor's Network project, studies of the existing regulation of social networks in Eastern and Central European countries in English will be available to interested readers. The project will therefore enable the general and specialized public to become acquainted with the existing regulations and future trends in this field. It will also contribute to the debate on pursuing a fragile balance between freedom of expression and the ideal of a secure and inclusive internet.

Control over content that reaches users of online services is never value-neutral. We can be skeptical about the content moderation carried out by social networks – as market players, they are primarily interested in profits. The recent revelations by the whistleblower Frances Haugen, who spoke about the unethical practices of Facebook and Instagram, are just another signal that social networks are concentrating a lot of power with potentially detrimental consequences for

society.¹⁴ Self-regulation is certainly not an appropriate solution to the problems associated with high concentration of power in companies that seemingly offer social networking services for free. Of course, many other issues remain open, e.g.: can we trust states and their organizations to be able to ensure effective and appropriate regulation of such enormous and complex systems?; how to ensure democratic oversight over social networks, establish efficient regulatory regimes, and how to involve users in these processes?; how to establish the elusive balance between freedom of expression and other human rights?... Considering that credible information and tolerant discussions are a precondition for a democratic society, these issues cannot be avoided in the future.

¹⁴ Dan Milmo: Frances Haugen: 'I Never Wanted to Be a Whistleblower. But Lives Were in Danger,' in: *The Guardian*, 24 October 2021; <<https://www.theguardian.com/technology/2021/oct/24/frances-haugen-i-never-wanted-to-be-a-whistleblower-but-lives-were-in-danger>> (25. 11. 2021).