LESSONS FROM THE RUSSIAN-UKRAINIAN WAR FOR BIG TECH'S MODERATION MECHANISMS: COMBATING INCITEMENT TO VIOLENCE

by Oleksandr Vasylenko

LLM Final Thesis

International Business Law - 2021/22

Supervisor: Maria José Schmidt-Kessen

Central European University - Private University

2022

The following thesis is dedicated to the topic of digital incitement to violence and aims to find ways of combating it by the actors of internet governance – states, social networks, social sector, and users. Through the use of doctrinal and interdisciplinary methodologies, the manner of social networks' moderation activities is analyzed in the framework of hateful speech towards the Ukrainian population in the context of the Russian-Ukrainian war. The focus is put on the jurisdictions and social networks, which are most actively triggered in the course of this war. Therefore, legislation of the EU, Ukraine, and Russia is scrutinized, as well as the actions of Facebook, Instagram, Twitter, YouTube, TikTok, and Telegram. The analysis of theories of internet governance and intermediary liability altogether with legislative provisions that regulate the sphere of incitement to violence's prohibition allows for assessing the social networks' modus operandi concerning the policing of this crime. Regarding the findings, the most fruitful way of counteraction to online hateful content is the cooperation of online content governance stakeholders to implement the mechanism of intermediary liability. The approaches of European legislators are the most prominent examples of such cooperation, which resulted in the social networks' greatest compliance in Europe. Although social networks have been taking steps to limit the presence of illegal incitements, it is still possible to find hateful content on each platform. Therefore, actors of internet governance should take proactive actions to combat online injustices in the course of the Russian-Ukrainian war and other possible scenarios.

ACKNOWLEDGEMENTS

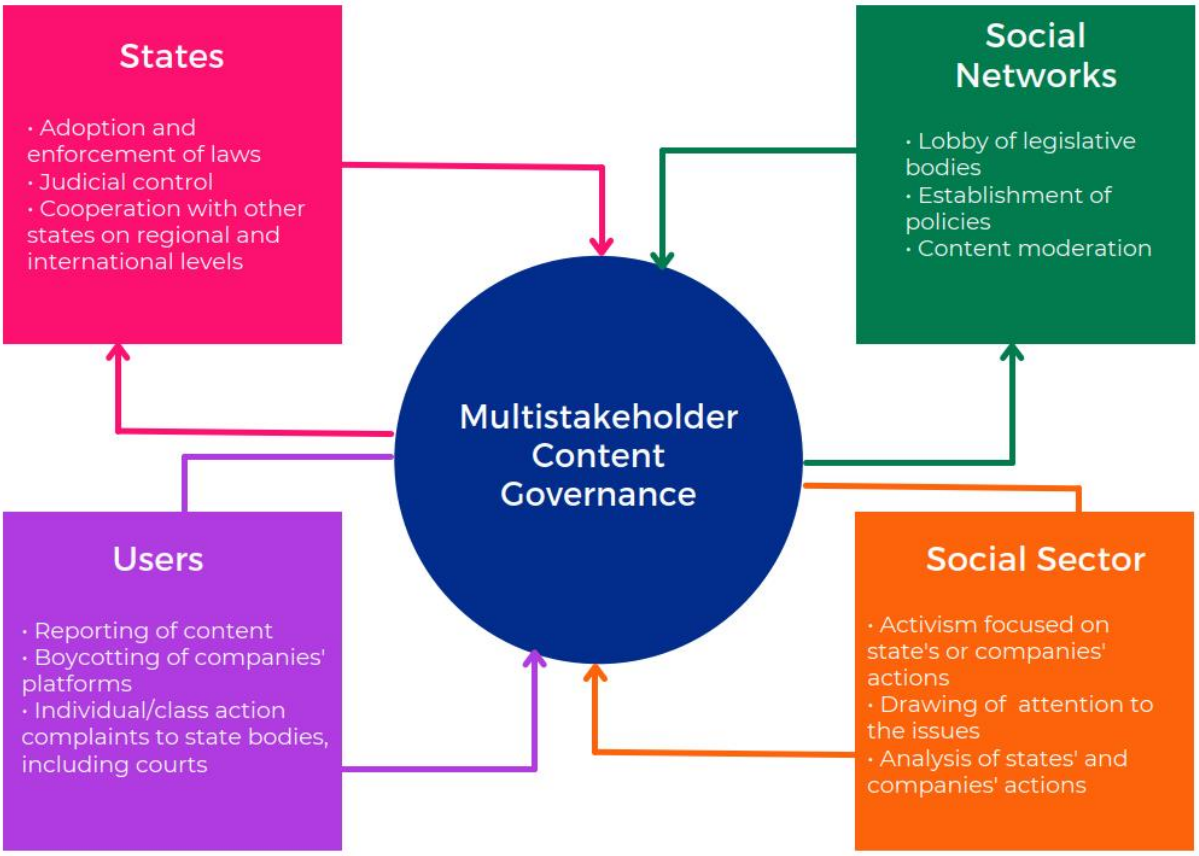
I want to express my deepest gratitude to my supervisor and lecturers for the masterful guidance over the last academic year and for the tons of precious knowledge that have been broadening the scope of my understanding of the law and its value. Your work had an immense influence on this thesis and will have a lasting impact on my future professional and academic endeavors.

Special thanks go to my family and friends for providing endless inspiration and support on this journey. Your love, warmth, and understanding, have always been the inalienable attributes of my accomplishments.

Besides, I would like to acknowledge the contribution of the Armed Forces of Ukraine throughout the war, which is yet to see its end. Your valor and sacrifices maintain the peace in the rest of Europe and save thousands of lives each day.

TABLE OF CONTENTS

LIST OF FIGURES



LIST OF ABBREVIATIONS

CPA	Austrian Federal law on measures to protect users on communication platforms (Communication Platforms Act - KoPl-G)
CSR	Corporate social responsibility
DSA	Digital Services Act
EU	European Union
ISPs	Internet service providers
NetzDG	German Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)
OHCHR	Office of the United Nations High Commissioner for Human Rights
UN	United Nations
UNGP	United Nations Guiding Principles on Business and Human Rights

INTRODUCTION

On 24 February 2022, the world witnessed a new chapter in the history of the Russian-Ukrainian war that has been tormenting millions of Ukrainians for more than eight years. The scale of humanitarian and economic losses is still to be established, yet the preliminary estimations evaluate the damages as around $600 billion.[1] The horrifying images of Ukrainian civilians being blatantly killed by Russian occupying forces made the global community wonder about the efficiency of international instruments and institutions, such as the UN, OSCE, and many others. While the conclusion of the war is yet to come, at the moment of this work's writing process, there are already lively discussions regarding the legal aftermath of Russian aggression.[2] One of the platforms, which could address the horrendous ongoing crimes, is the ICC, which has already opened an investigation of the current fateful actions.[3]

While the outcome of the war seems blurry and will take an unknown amount of time to conclude, the prerequisites of the war and deployed instruments are already available for analysis. Due to the modern nature of the war, since 2014 it has been getting a connotation of "hybrid", which was destined to underline the mixture of various types of warfare, apart from conventional ones. Thus, for eight years the elements of lawfare, political warfare, cyber warfare, and other types of tools have been utilized to achieve victory by both sides.[4] In this sense, the digital sphere became one of the main battlefields for the clash over the minds of Internet users from both Ukraine and Russia and the sympathies of users from around the globe. The attacks on banking infrastructure through the use of malware,[5] armies of bots storming the

[1] Madeline Halpert, 'Russia's Invasion Has Cost Ukraine Up To $600 Billion, Study Suggests' *Forbes* (2022) <https://www.forbes.com/sites/madelinehalpert/2022/05/04/russias-invasion-has-cost-ukraine-up-to-600-billion-study-suggests/> accessed 31 August 2022.

[2] Financial Times, 'Holding Russia To Account For War Crimes' (2022) <https://www.ft.com/content/aacc2e1d-d450-4345-bedf-2cfb1af60e8a> accessed 31 August 2022; Kevin Jon Heller, 'Creating A Special Tribunal For Aggression Against Ukraine Is A Bad Idea' <https://opiniojuris.org/2022/03/07/creating-a-special-tribunal-for-aggression-against-ukraine-is-a-bad-idea/ (Last accessed: 25 August 2022)> accessed 31 August 2022; Shweta Desai, 'PACE Calls For International Tribunal To Probe Russian War Crimes In Ukraine' <https://www.aa.com.tr/en/europe/pace-calls-for-international-tribunal-to-probe-russian-war-crimes-in-ukraine/2575904 (Last accessed: 25 August 2022)> accessed 31 August 2022.

[3] 'Ukraine: Situation In Ukraine - ICC-01/22' (*International Criminal Court*, 2022) <https://www.icc-cpi.int/ukraine> accessed 31 August 2022.

[4] Frank Hofmann, 'The Hybrid War That Began Before Russia Invaded Ukraine' *DW* (2022) <https://www.dw.com/en/hybrid-war-in-ukraine-began-before-russian-invasion/a-60914988> accessed 31 August 2022.

[5] 'Alert (AA22-110A)' (*Cybersecurity & Infrastracture Security Agency*) <https://www.cisa.gov/uscert/ncas/alerts/aa22-110a> accessed 31 August 2022.

comments sections on social networks,[6] advertising campaigns heavily dependent on personal data to influence the course of elections,[7] and the overwhelming amount of fake news sowing disbelief across digital platforms[8] – all these methods have been actively deployed even outside the boundaries of the two states at war.

A special role in digital warfare is devoted to the dissemination of various hate speech massages, including incitements to violence, aimed at multiple audiences and able to shape up a radicalized worldview, which can lead to the active participation in the acts of violence, toleration of them, and even ignorance to the victims. The malicious use of incitement to violence has been a reliable companion of state-led atrocities and policies that led to the genocide of thousands of people, with Goebbels-led Nazi propaganda as one of the most prominent examples.[9] Despite the lessons of the Nuremberg Trials and the principles, established in the *Streicher* case,[10] humanity witnessed the hateful use of mass media on numerous occasions, from the radio broadcasts that contributed to the Tutsi genocide in Rwanda,[11] to the calls for violent actions against the Rohingya population via Facebook.[12] Unfortunately, history repeats itself in the ongoing war between Russia and Ukraine, where the use of technologies and hateful messages reaches unprecedented extents.

Judging from the actions of major social media platforms, hate speech and incitement to violent actions are not left unnoticed, since some proactive combating actions are done by

[6] BBC, 'How Russian bots appear in your timeline' (2017) <https://www.bbc.com/news/technology-41982569> accessed 31 August 2022.

[7] Philip Bump, 'What Data On More Than 3,500 Russian Facebook Ads Reveals About The Interference Effort' *The Washington Post* (2018) <https://www.washingtonpost.com/news/politics/wp/2018/05/10/what-data-on-more-than-3500-russian-facebook-ads-reveals-about-the-interference-effort/> accessed 31 August 2022.

[8] Kathrin Wesolowski, 'Fact Check: Fake News Thrives Amid Russia-Ukraine War' *DW* (2022) <https://www.dw.com/en/fact-check-fake-news-thrives-amid-russia-ukraine-war/a-61477502> accessed 31 August 2022.

[9] Heidi Tworek, 'A Lesson From 1930S Germany: Beware State Control Of Social Media' *The Atlantic* (2019) <https://www.theatlantic.com/international/archive/2019/05/germany-war-radio-social-media/590149/> accessed 31 August 2022; The New Yorker, 'Copenhagen, Speech, And Violence' (2015) <https://www.newyorker.com/news/news-desk/copenhagen-speech-violence> accessed 31 August 2022.

[10] 'Nuremberg Trial Proceedings Vol. 12' (*The Avalon Project*) <https://avalon.law.yale.edu/imt/04-29-46.asp> accessed 31 August 2022; Wibke Kristin Timmermann, 'Incitement In International Criminal Law' (2006) 88 International Review of the Red Cross, p. 827.

[11] Ibid., p. 841; *Prosecutor v. Ruggiu* (Judgement and Sentence) ICTR-97-32-I (1 June 2000), paragraphs 16, 22.

[12] Rodion Ebbighausen, 'Inciting Hatred Against Rohingya On Social Media' *DW* (2018) <https://www.dw.com/en/inciting-hatred-against-rohingya-on-social-media/a-45225962> accessed 31 August 2022.

their moderation teams. However, the approaches of social platforms have not been identical and greatly vary in efficiency. Therefore, there is an issue of introducing a unified approach that would be an efficient deterrent from incitement to violence. And, once such an approach is established, there should be a way to make it obligatory for the platforms.

To come up with an efficient one-size-fits-all approach, it is necessary to understand the nuances of power interplay on the Web and what would be the most beneficial strategy to use in this interplay in the context of combating online incitements to violence. In this regard, it would be beneficial to understand how states and intermediaries influence one another in the spectrum of incitement to violence on both national and international levels, and how users and social sector can play a role in this interaction. Consequently, it is useful to look for the most viable mechanism, which can be deployed by the aforementioned actors. The aforementioned information will allow us to answer the major questions of this work: 1) how the existing legislation deals with digital incitement to violence in the context of the Russian-Ukrainian war, 2) in which manner social networks respond to such legislation, and 3) how can actors of internet governance counteract to ongoing and future digital incitements to violence?

METHODOLOGY

The following work will deploy both doctrinal and interdisciplinary methodologies to achieve the aforementioned aims. As for doctrinal methodology, I will use legal theory research to analyze theories and strategies applicable to digital incitement to violence, particularly, those of internet governance and intermediary liability. Moreover, expository research will be deployed for the assessment of existing legal norms on international, regional, and national levels that relate to the respect of human rights by businesses, freedom of expression, incitement to violence, and the balancing of the latter two. These acts will provide the legal framework for the assessment of digital incitement to violence in the course of the Russian-Ukrainian war and prerequisites for the actions of social networks. The socio-legal research will be used to explain the interplay of actors in the digital sphere over the topic of incitement to violence. In particular, I will use the interdisciplinary theory of internet governance from the domains of law, social and Internet studies to scrutinize the reception of legal norms by social networks and the legality of their moderation activities over the content, which relates to the Russian-Ukrainian war and can be qualified as incitement to genocide. For this purpose, I will also use empirical evidence that shows how social networks react to the reports of users. This information will help to analyze the relationships between the four most prominent actors in the scheme of multistakeholder digital content governance – states, social networks, social sector, and users. However, as states and social networks play the most prominent role in the implementation of online actions, the major emphasis will be put on them.

The current Russian-Ukrainian war was chosen as a framework for this work not only because of its grave nature and importance to the forthcoming changes in international political and legal states of affairs but also because of the abundance of examples it can provide for the topic of incitement to violence. Despite the previous and ongoing instances of such crimes (Nazi propaganda against Jewish people, Rwandan crimes against Tutsi, incitements targeted at Rohingya in Myanmar), the Russian-Ukrainian war gave rise to the unseen volume of hateful content, which can be used as a litmus test to assess the efficiency of current responses to incitement to violence by the actors of internet governance. Hopefully, this work will not only contribute to the study of the topic of internet crimes and their prevention, but will also provide the basis to vitiate negative effects of incitement to violence and hate speech in the course of the current war, and the future possible instances of digital incitements.

As to the selection of jurisdictions, apart from the international framework of the United Nations (further – "UN") and its bodies, I will focus on regional and national regimes that have a connection to the issue of digital incitement to violence against Ukrainians. Since the spillover effects of the Russian-Ukrainian war concern not only the primary parties of the war, the accent is also put on the European Union (further – "EU") itself, and some of its Member States that take important decisions as to the effects of the war in the digital sphere and have legislation that directly addresses the issues of online incitement to violence, such as Germany and Austria. Moreover, the residents of these areas can freely access hateful content against Ukrainians, as the social networks rarely have physical borders, and the number of Ukrainians in the EU is quite prominent as well.

Moving to the selection of social networks, it is made according to the statistics of their popularity among the residents of Ukraine, Russia, and the EU, their use to highlight the news and events relating to the war, and the potential to influence public opinion.[13] Therefore, social networks in scope are Meta's Facebook and Instagram, Twitter, YouTube, TikTok, and Telegram. To assess their moderation activities, I will analyze their policies, specific actions during the course of the Russian-Ukrainian war and user-reporting functions. Furthermore, I will conduct a superficial search for infringing content through the prism of incitement to violence's definitions from Section II.B.2 of this work. For the sake of this experiment, any content that falls into international, national or social networks' definition of incitement to violence will be qualified as infringing. The infringing content will be reported to the moderation teams with the further analysis of their reaction. All the posts will be gathered in the specific Google Drive folder in case they get deleted.[14] This experiment will not be fully representative as to the moderation activities of platforms apart from the described situations. However, they may help to illustrate the reporting mechanisms, assessed in this work.

[13] Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022; 'Penetration Of Selected Social Media Platforms In Ukraine And Russia As Of November 2021' (*Statista*) <https://www.statista.com/statistics/1308258/social-media-penetration-ukraine-russia/> accessed 31 August 2022.

[14] 'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022.

Naturally, the given response will not be the most comprehensive due to the necessity of deep interdisciplinary analysis, which should include the evaluation of large quantities of data, the broader assessment of legislative answers across the globe and full cooperation of social networks, including the disclosure of monitoring activities and reaction to the user-reports. In my thesis, I will not focus deeply on legislation outside of chosen jurisdictions or go out of the framework of Russian-Ukrainian war. Moreover, I will not conduct profound sociological experiments to fully assess the modus operandi of social networks. Such limitations are made for the sake of coherence and clear causality in the presented solutions. Considering the research restraints, the findings of this work can serve as an initial guidance to the key players in the field of content moderation and are free to be supplemented by the results of related research activities. In the end, the panacea to the existing issue can be found only in a cooperative mode that would only benefit from the interdisciplinary approach.

Finally, before we proceed to the substantive part, it is necessary to establish the prime ethical pillars on which this work stands. First of all, honesty and transparency are the main principles that guided the way to the findings of this thesis. I aimed to present such findings with the full and unequivocal representation without misleading information. The empirical part was collected with the utmost respect for the people's privacy and dignity. The only data from social networks included in this thesis is the one available to the general public and is presented as it is. Apart from the position of a scholar, I collected information as a user of social networks. This approach helped me to provide recommendations not only from the field of legal theory but also from the user's experience. In the end, my work was directed by the paramount respect for human rights and the authority of international standards on incitement to violence, which motivated me to find ways to contribute to the protection of the actual and potential victims of hateful speech.

With that being said, let us dive into the domain of internet governance over incitement to violence.

CHAPTER I: LIVING IN THE ERA OF ONLINE INCITEMENT TO VIOLENCE: THE INTERPLAY OF INTERNET STAKEHOLDERS IN THE CONTEXT OF INTERMEDIARY LIABILITY

The increasing role of technology in people's lives has served as a blessing for fast learners and a source of stress for those, unwilling or unable to adapt to the changing circumstances. Analyzing the lessons of the past few decades, the rapid digitalization of communicational and commercial realms of human interaction was a massive technological leap and still has a rocketing speed of development, which makes it hard not only for usual people but also for usual societal institutions to adapt to them adequately. Since the days of its creation, the World Wide Web has received the dosage of polarized views ranging from its appreciation as a technocratic state of the art to the dreadful comments regarding its fatal influence on the usual human communication.[15] Acknowledging the massive comforts the Internet has brought to our society, the beneficial nature of its use highly depends on the behavior of actors connected to it, and the online climate, crystallized as a result. In this sense, such a common social network as Facebook can both serve as the headquarters for cat-lovers, and the major dissemination base for incitement to violence against Rohingya people.[16] Thus, to understand the factors, which influence the Web on a daily basis, it would be useful to identify the factors, which formed the present *status quo*. In this sense, special attention will be given to the legal definition of social networks, and to theoretical and strategic approaches that can be applied to it in the context of online crimes counteraction. To satisfy the second need we will focus on multistakeholder internet governance and intermediary liability.

I.A. The legal features of social networks

A new chapter in the Internet's history was initiated, when the concept of Web 2.0 got its practical implications in the form of participatory web pages, where users were actively engaging in the processes of creation and sharing of content, instead of data consumption from

[15] European Parliamentary Research Service, 'Potentially Negative Effects Of Internet Use' (European Parliament 2022); Janet Abbate, 'The Internet: Global Evolution And Challenges' <https://www.bbvaopenmind.com/en/books/frontiers-of-knowledge/> accessed 31 August 2022.

[16] Dan Milmo, 'Rohingya Sue Facebook For £150Bn Over Myanmar Genocide' *The Guardian* (2021) <https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence> accessed 31 August 2022.

a unified source.[17] The emergence of Web 2.0 practically enabled the active dissemination of illegal content and created the need for active policing actions on behalf of platforms moderation teams. However, not all Web 2.0 platforms pose interest for the legislators. Therefore, it is necessary to identify whether the networks, most actively used in the context of Russian-Ukrainian war are qualified as online platforms under the relevant legislation.

Acknowledging the subject matter of this work, which is incitement to violence, it would be most useful to turn to the definitions, present in legislative acts, which directly address this issue, namely the German Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) (further – "NetzDG")[18] and Austrian Federal law on measures to protect users on communication platforms (Communication Platforms Act - KoPl-G) (further – "CPA").[19] The acts define the scope of actors, responsible for the fulfillment of monitoring and acting obligations, diversifying between regular communication platforms (social networks) and video sharing platforms. The social networks are described as "telemedia service providers who operate platforms on the Internet with the intention of making a profit, which are intended for users to share any content with other users or make it accessible to the public" in NetzDG[20] and as "an information society service where the main purpose or essential function is to enable the exchange of communications or performances containing ideas, whether spoken, written, audio or visual, by means of mass dissemination, between users and a wider range of other users" in CPA.[21] As to the definition of video sharing platforms, NetzDG defines it as "telemedia where the main purpose or an essential function is to make broadcasts or user-generated videos available to the general public, for which the service provider has no editorial responsibility, with the service provider determining the organization of the broadcasts or user-generated videos, including by automatic means",[22] and CPA stipulates it as "a service in which the main purpose or a separable part of the service or an essential function of the

[17] David Wilson and others, 'Web 2.0: A Definition, Literature Review, And Directions For Future Research' [2011] AMCIS 2011 Proceedings - All Submissions.

[18] Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG) 2017.

[19] Bundesgesetz über Maßnahmen zum Schutz der Nutzer auf Kommunikationsplattformen (Kommunikationsplattformen-Gesetz – KoPl-G) 2020.

[20] NetzDG, § 1 Scope of Application, Definitions.

[21] CPA, § 2(2).

[22] NetzDG, § 3d Definitions for video sharing platform services.

service consists in broadcasting or, user-generated videos or both, for which the platform provider bears no editorial responsibility, to the general public."[23] According to these definitions, social networks include such companies as Facebook, Instagram, Twitter, and Telegram,[24] whilst video sharing platforms include YouTube and TikTok. Consequently, the aforementioned definitions outline the most prominent features of content platforms and are applicable to the networks, analyzed in this work.

I.B. Multistakeholder governance as the instrument of online changes

When we think about the Internet, we may vastly simplify the interactions, which are occurring in the field of decision-making. However, the Web is a place for the plethora of constant interactions among online actors who have the potential to influence each other. One of the best frameworks that describe the interplay among online actors is the theory of Internet governance. It should be noted that the views on Internet governance have not yet been standardized, which is explained by the number of actors with different professional backgrounds and ideological standpoints who study this matter.[25] As was noted by Ziewitz and Pentzold, there are some disputes as to the relevant players, who have the authority over the online field, and the areas where this authority is deployed as well.[26] For instance, while Johnson underlines the importance of the users' role in the creation of "social order online", Braman and Palfrey focus on more stringent policy and regulations-based approaches.[27] Another area of inconsistency is the subject matter of Internet governance. While Benkler describes it as a multilayered governance system, which includes physical infrastructure, code or logic, and content layers,[28] DeNardis and Kurbalija take a more abstractive stance, claiming that Internet governance relates to "policy and technical coordination issues related to the

[23] CPA, § 2(12).

[24] 'How Many People Use Telegram In 2022? 55 Telegram Stats' (*Backlinko*) <https://backlinko.com/telegram-users> accessed 31 August 2022; Archyworldys, 'Telegram Should Adhere To The Netzdg' (2021) <https://www.archyworldys.com/telegram-should-adhere-to-the-netzdg/> accessed 31 August 2022.

[25] Malte Ziewitz and Christian Pentzold, 'In Search Of Internet Governance: Performing Order In Digitally Networked Environments' (2014) 16 New Media & Society.

[26] Ibid.

[27] Sandra Braman, 'Internet Policy', *The Handbook of Internet Studies* (Wiley-Blackwell 2010), pp. 137-167; John Palfrey, 'Four Phases Of Internet Regulation' (2010) 77 Social Research: An International Quarterly, pp. 981-996.

[28] Yochai Benkler, 'From Consumers To Users: Shifting The Deeper Structures Of Regulation Towards Sustainable Commons And User Access' (2000) 51 Federal Communications Law Journal.

exchange of information over the Internet"[29] and includes a much broader variety of issues than the "infrastructural aspects."[30]

The history of Internet governance demonstrates that multiple organizations exerted a certain degree of authority over various aspects of the Web[31], but human rights are rarely the primary subject matter of their operations. One of the first international efforts that did focus on human rights-related issues was the Working Group on Internet Governance, initiated by the UN Secretary-General to elaborate on such topics as the definition of Internet governance, identification of relevant public issues relating to the Internet, and the allocation of the roles and responsibilities to the governments before the 2005 World Summit on the Information Society in Tunis.[32] One of the Summits results was the actual definition of Internet governance, which stated that it is "the development and application by Governments, the private sector, and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programs that shape the evolution and use of the Internet."[33] Prominently, by naming several relevant actors, the definition highlighted the multistakeholder nature of online governance.

The concept of multistakeholder internet governance has been rising to prominence as one of the most important elements of the Internet's development and unified answer to legal, and sociopolitical issues, related to the Web.[34] Multistakeholder approach has also found its

[29] Laura DeNardis, 'The Emerging Field Of Internet Governance' [2010] Yale Information Society Project.

[30] Jovan Kurbalija, *An Introduction To Internet Governance* (5th edn, DiploFoundation 2012).

[31] 'Who We Are' (*IETF*) <https://www.ietf.org/about/who/> accessed 31 August 2022; 'Corporate Documents' (*ARIN*) <https://www.arin.net/about/corporate/documents/> accessed 31 August 2022; 'What We Do' (*RIPE NCC*) <https://www.ripe.net/about-us/what-we-do> accessed 31 August 2022; 'APNIC Serves The Asia Pacific Region' (*APNIC*) <https://www.apnic.net/about-apnic/organization/apnic-region/> accessed 31 August 2022; 'Acerca De LACNIC' (*LACNIC*) <https://www.lacnic.net/966/1/lacnic/acerca-de-lacnic> accessed 31 August 2022; 'AFRINIC-31' (*AFRINIC*) <https://meeting.afrinic.net/afrinic-31/en/about/afrinic-31> accessed 31 August 2022.

[32] Working Group on Internet Governance, 'Report Of The Working Group On Internet Governance' (2005) <http://www.wgig.org/docs/WGIGREPORT.pdf> accessed 31 August 2022, p. 4.

[33] Ibid.

[34] UNESCO, 'Keystones To Foster Inclusive Knowledge Societies: Access To Information And Knowledge, Freedom Of Expression, Privacy And Ethics On A Global Internet' (UNESCO 2015) <https://unesdoc.unesco.org/ark:/48223/pf0000232563> accessed 31 August 2022; Council of Europe, Convention on Cybercrime (Budapest, 23 November 2001) European Treaty Series - No. 185; African Union, Convention on Cyber Security and Personal Data Protection (Malabo, 27 June 2014); 3. Proposal for a Regulation on Digital Markets Act (Digital Markets Act) 2020; 'The Digital Markets Act: Ensuring Fair And Open Digital Markets' (*European Commission*) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en> accessed 31 August 2022, Preamble (105); Proposal for a Regulation on a Single Market For Digital Services (Digital Services Act) 2020, Preamble

place in the EU legislation, particularly, in Digital Markets Act,[35] and Digital Services Act (further – "DSA"),[36] with the latter playing the most significant role in combating incitement to violence, which will be studied in the following sections. Moreover, the Internet Governance Forum has already analyzed the issue of multistakeholder initiatives in content governance.[37]

Building upon the theoretical standpoints, international and regional practices in multistakeholder internet governance within the content-oriented matter, the definition that is narrowly tailored to the combating of digital incitement to violence is proposed. Consequently, in this work, multistakeholder content governance is to be understood as the constant elaboration and implementation of Internet-related principles, rules, mechanisms, and procedures in the field of content sharing and moderation by states, social networks, social sector (activists and academia), and users with interdependence between each other. Moreover, the following scheme (Figure 1) would sufficiently mirror the aforementioned definition.

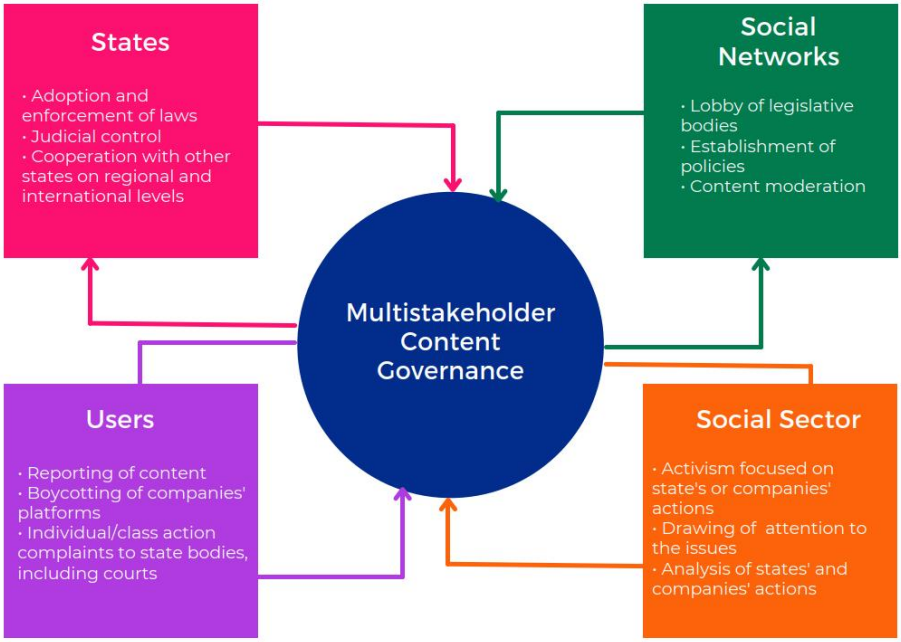


Figure 1

(70), Articles 35(2), 41(1)(e); 'The Digital Services Act: Ensuring A Safe And Accountable Online Environment' (*European Commission*) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en> accessed 31 August 2022.

[35] Digital Markets Act, Preamble (105).

[36] DSA, Preamble (70), Articles 35(2), 41(1)(e).

[37] 'IGF 2021 WS #57 Multistakeholder Initiatives In Content Governance' (*IGF Internet Governance Forum*) <https://www.intgovforum.org/multilingual/content/igf-2021-ws-57-multistakeholder-initiatives-in-content-governance> accessed 31 August 2022.

It is necessary to keep this definition and scheme in mind, as we move forward into the realm of digital incitement to violence. The model of multistakeholder internet governance has the potential of becoming a groundbreaking tool for dealing with online crimes. However, the instrument itself will not bring the awaited result without an effective strategy for its use. And the answer to this riddle is the doctrine of intermediary liability.

I.C. Intermediary liability as the strategy against incitement to violence

The doctrine of intermediary liability refers to the approach, in which the intermediary (usually the ISPs and social platforms) is liable for the violations that occur on its website, or are created by its users and are not policed by the intermediary as the law requires.[38] The doctrine has found its application in both legislative and judicial dimensions.

Concerning legislation, intermediary liability has been one of the major responses to hate speech, including incitements to violence, being present in such legislative acts, as NetzDG, CPA, and the forthcoming DSA. These laws share similar features, which are inalienable for successful implementation. One of them is the imposition of certain obligations, which incentivizes ISPs to take action against illegal content and its publishers.[39] The second feature is the deployment of a specific reactive mechanism, which provides intermediaries with certain algorithms of actions, when illegal content is spotted, with notice-and-takedown being the most prominent example.[40] Furthermore, there is a limitation of ISPs' liability, which usually refers to the absence of knowledge about illegal materials or the actual implementation of the prescribed actions.[41]

Such qualities of intermediary liability as the limited allocation of a burden on ISPs and the imposition of moderation requirements make them a perfect embodiment of the multistakeholder approach to the apprehension of online incitements to violence. However, intermediary liability can be implemented in a variety of ways, and can even lead to the

[38] Alex Comninos, 'The Liability Of Internet Intermediaries In Nigeria, Kenya, South Africa And Uganda: An Uncertain Terrain' [2012] Intermediary Liability in Africa Research Papers, p. 6.

[39] Ashley Johnson and Daniel Castro, 'How Other Countries Have Dealt With Intermediary Liability' *Information Technology & Innovation Foundation* (2021) <https://itif.org/publications/2021/02/22/how-other-countries-have-dealt-intermediary-liability/> accessed 31 August 2022.

[40] Ibid.

[41] Ibid.

violation of fundamental constitutional rights, as in the example of "Avia" law.[42] In this regard, the deployment of intermediary liability is always associated with certain risks and should satisfy numerous criteria, including legality, legitimacy, necessity, and respect for privacy.[43] Therefore, to identify the best approaches for combating incitement to violence, we should pinpoint the necessary international legislative provisions and analyze their implementation by such actors of internet governance as states and social networks. With this in mind, we shall move to the analysis of the existing international obligations in our subject matter and the stakeholders' compliance with them.

[42] 'France: Constitutional Council Declares French Hate Speech 'Avia' Law Unconstitutional' <https://www.article19.org/resources/france-constitutional-council-declares-french-hate-speech-avialaw-unconstitutional/> accessed 31 August 2022.

[43] Ibid.; Global Network Initiative, 'Content Regulation And Human Rights' (Global Network Initiative 2020) <https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf> accessed 31 August 2022; *Magyar Tartalomszolgáltatók Egyesülete and Indexhu Zrt v Hungary* App no 22947/13 (ECtHR, 2 February 2016); Michael Karanicolas, 'Newly Published Citizens Protection (Against Online Harm) Rules Are A Disaster For Freedom Of Expression In Pakistan' <https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-information/wiii-blog/newly-published-citizens-protection-against-online-harm-rules-are-disaster-freedom-expression> accessed 31 August 2022.

II. IMPERIUM, MEDIATOR, POPULUS: STANDARDS FOR HUMAN RIGHTS PROTECTION IN THE WEB AND THEIR IMPLEMENTATION

With respect to the functioning of human rights in the online sphere, the impact of digital businesses has additional importance due to their international character and position of power to dictate the rules for specific networks.[44] In particular, social networks can influence state policy making,[45] change the way governments and legislators interact with the citizens,[46] shape the political debate[47] and even be a law enforcer in certain situations.[48] To ensure human rights compliance of these actors, the international community managed to issue the United Nations Guiding Principles on Business and Human Rights (further – "UNGP").[49] Moreover, in a similar cooperative mode, the principles of freedom of expression and incitement to violence were established. Due to the absence of strong enforcement mechanisms, there is a potential for uneven implementation of the aforementioned rules that led to the dramatic difference in legal regimes for the balancing of freedom of expression and incitement to violence.[50] To understand the nuances of this balancing, in this section, I will focus on relevant international rules, which relate to online incitement to violence and their application by states and private companies. Starting from the general obligation for states and businesses to respect human rights enshrined in UNGP, the narration will shift to subsequent rules that govern the sphere of freedom of expression, incitement to violence, and the balancing of the latter two. This information will

[44] Sian Jones, 'The Social Dilemma And The Human Rights Risks Of Big Tech' <https://www.humanrightspulse.com/mastercontentblog/the-social-dilemma-and-the-human-rights-risks-of-big-tech> accessed 31 August 2022; Deborah Brown, 'Big Tech's Heavy Hand Around The Globe' <https://www.hrw.org/news/2020/09/08/big-techs-heavy-hand-around-globe> accessed 31 August 2022; 'Tech Giants And Human Rights: Investor Expectations' <https://www.humanrights.dk/sites/humanrights.dk/files/media/document/Tech%20giants%20and%20human%20rights_2021.pdf> accessed 31 August 2022.

[45] Marco Battaglini and Eleonora Patacchini, 'Social Networks In Policy Making' (2019) 11 Annual Review of Economics.

[46] Shelby Sklar, 'The Impact Of Social Media On The Legislative Process: How The Speech Or Debate Clause Could Be Interpreted' (2015) 10 Northwestern Journal of Law and Social Policy.

[47] José Luis Vargas Valdez, 'Study on the Role of Social Media and the Internet in Democratic Development' [2018] Venice Commission.

[48] Katharina Kaesling, 'Privatising Law Enforcement In Social Networks: A Comparative Model Analysis' (2018) 11 Erasmus Law Review.

[49] Rachel Davis, 'The UN Guiding Principles On Business And Human Rights And Conflict-Affected Areas: State Obligations And Business Responsibilities' (2012) 94 International Review of the Red Cross.

[50] Dominika Bychawska-Siniarska, *Protecting The Right To Freedom Of Expression Under The European Convention On Human Rights* (Council of Europe 2017).

give an understanding of the prerequisites for business navigation during the ongoing Russian-Ukrainian war, and whether their actions correspond to international requirements.

II.A. The legal prerequisites of obligation to protect: UNGP and its unequal implementation

Nowadays, the idea of respect for human rights is an inalienable part of international public law and receives unanimous support in the acts of such international organizations as the UN,[51] the Council of Europe,[52] and many other regional and global bodies.[53] Throughout the development of human rights doctrines, their protection and implementation were predominantly based on states.[54] However, the growing influence of the private sphere in the form of enterprises and the growing number of human rights violations conducted by them led to a slight change of focus and made legal scholars scrutinize the question of business accountability.[55] Over a long process of negotiations and drafting, the UNGP were finally introduced in 2011, which marked the creation of an authoritative source in the field of states-business-human rights nexus.[56] The nature of this document is two-sided since it contains both the obligations for states and businesses, emphasizing their ultimate role in human rights protection.[57] One of the greatest UNGP's achievements is the introduction of a state-business nexus that accounts for the state protection and promotion of human rights compliance by business. In addition to the oversight and regulatory obligations, imposed on states, UNGP enlists several requirements focused on businesses. In particular, there is a general obligation

[51] Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR).

[52] Council of Europe, European Convention on Human Rights (Rome, 4 November 1950) (ECHR).

[53] 'Conventions And Recommendations' (*International Labor Organization*) <https://www.ilo.org/global/standards/introduction-to-international-labour-standards/conventions-and-recommendations/lang--en/index.htm> accessed 31 August 2022; 'The Role Of The High Commissioner For Human Rights' (*OHCHR*) <https://www.ohchr.org/en/about-us/high-commissioner> accessed 31 August 2022; 'What We Do' (*UNICEF*) <https://www.unicef.org/what-we-do> accessed 31 August 2022; 'Mandate Of The Commission' (*African Commission on Human and Peoples' Rights*) <https://www.achpr.org/mandateofthecommission> accessed 31 August 2022; 'What We Do' (*ASEAN*) <https://asean.org/what-we-do/> accessed 31 August 2022; 'What We Do' (*OAS*) <https://www.oas.org/en/about/what_we_do.asp> accessed 31 August 2022.

[54] 'The Foundation Of International Human Rights Law' (*UN*) <https://www.un.org/en/about-us/udhr/foundation-of-international-human-rights-law> accessed 31 August 2022.

[55] Erika George, *Incorporating Rights* (5th edn, Oxford University Press 2021), pp. 65-96.

[56] Ibid.

[57] OHCHR 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04.

to respect human rights through avoidance of infringements and prevention of potential violations linked to their activities.[58] Regarding the means, through which this obligation can be met, UNGP proposes a three-way algorithm, which includes 1) policy commitment, 2) human rights due diligence, and 3) remediation.[59]

Ultimately, the UNGP is a highly important document that gave rise to the closer adherence of businesses to internationally recognized human rights standards. The idea of binary responsibility of both states and enterprises to protect human rights creates a solid framework for further cooperation between national authorities and private actors. However, this act was just an initial step toward business-centered obligations and did not stop the ongoing human rights violations. The reasons for the limited efficiency include the non-binding nature of this document and the absence of strict operational guidelines. Both factors contribute to the flexibility of the UNGP's implementation and give the actors some scope and freedom to tailor the general requirements to practical peculiarities. However, judging this document from an 11-year perspective, it failed to introduce a unified paradigm of state and business respect for human rights, which makes it necessary to analyze the legislative practices and business adherence connected to UNGP.

II.A.1. Regional and national UNGP implementation: from ignorance to full-scale CSR

The implementation of UNGP by states has been highly unequal, as some countries have taken certain steps toward corporate social responsibility (further – "CSR") even before the UNGP's entry into force, while the other states failed to produce at least a vague reminiscent of national actions plans.[60] The EU has been the most active promoter of CSR for the last two decades, arguing the importance of its legislative introduction in the 2001 green paper[61] and other documents, such as "A renewed EU strategy 2011–14 for Corporate Social Responsibility."[62] Moreover, the majority of EU members and such European countries, as the

[58] Ibid., Principles II(A)(11), (13)(a,b).

[59] Ibid., Principles II(B)(16-22).

[60] Laura Albareda, Josep M. Lozano and Tamyko Ysa, 'Public Policies On Corporate Social Responsibility: The Role Of Governments In Europe' (2007) 74 Journal of Business Ethics.

[61] European Commission, 'Green Paper Promoting A European Framework For Corporate Social Responsibility' (2001).

[62] European Commission, 'Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: A Renewed EU Strategy 2011–14 For Corporate Social Responsibility' (European Commission 2011); 'Corporate Social Responsibility &

UK, Norway, Switzerland, and Georgia managed to introduce comprehensive national action plans, which fulfill the postulates of UNGP.[63] Moreover, Portugal[64] and Ukraine[65] are in the final stages of national action plan development, which may be postponed in the latter country due to the hostile actions by Russia. As for the other continents, considerable progress was also achieved by the US, Commonwealth countries, and some Asian countries, such as Japan, Taiwan, and Thailand.[66] Moreover, the forthcoming EU legislation, such as DSA is also using the strategy of CSR, ensuring that big tech companies are obliged to conduct risk-assessment activities to delete illegal online content.[67] Overall, the majority of Western democracies, whose citizens are most exposed to incitement to violence in the context of the Russian-Ukrainian war tend to have national action plans, destined to ensure respect for human rights by the private sector. As to the parties of the war, only Ukraine has done considerable steps to promote human rights due diligence in business circles by introducing the National Strategy on Human Rights in 2021[68] and inviting other stakeholders to the assessment of the implementation process.[69]

II.A.2. Compliance of social networks with UNGP

Concerning the recognition of UNGP requirements by other stakeholders from the realm of content governance, major social networks generally recognize their human rights obligations and even create their own initiatives. For instance, in 2008 Google together with such technological giants as Microsoft and Yahoo created the Global Network Initiative, designed to promote human rights due diligence in the field of big tech with the focus on

---

Responsible Business Conduct' (*European Commission*) <https://single-market-economy.ec.europa.eu/industry/sustainability/corporate-social-responsibility-responsible-business-conduct_en> accessed 31 August 2022; 'Commission Staff Working Document - Corporate Social Responsibility, Responsible Business Conduct, And Business And Human Rights: Overview Of Progress' (*European Commission*, 2019) <https://ec.europa.eu/docsroom/documents/34482> accessed 31 August 2022.

[63] 'Countries' (*National Action Plans on Business and Human Rights*) <https://globalnaps.org/country/> accessed 31 August 2022.

[64] Ibid., Portugal.

[65] Ibid., Ukraine.

[66] Ibid.

[67] DSA, Article 26.

[68] President of Ukraine, 'Decree №119/2021 "On The National Strategy For Human Rights"' (President of Ukraine 2021).

[69] 'Countries' (*National Action Plans on Business and Human Rights*) <https://globalnaps.org/country/> accessed 31 August 2022.

freedom of expression.[70] Furthermore, such companies as Meta and Apple issued human rights-related documents, where they acknowledged the importance of UNGP and committed to them, stating that their approach is based on the mechanisms, established in the UN document.[71] Another approach, taken by Google and Twitter is the creation of dedicated pages on their websites that underline the companies' commitment to human rights and utmost respect to UNGP.[72] The other social media companies, which are actively used by the residents of Ukraine and Russia, TikTok and Telegram do not have any mentions of UNGP and human rights in their policies and websites.[73]

II.A.3. Public oversight over UNGP adherence

Finally, it is necessary to mention the role of users and the social sector in UNGP fulfillment. Whilst they are not explicitly mentioned in the UN document as the major stakeholders with the burden of implementation, their role is also highly important for the sake of the CSR of businesses, including social platforms. Due to the work of these two groups, such monitoring initiatives as Economy for the Common Good's Common Good Balance Sheet,[74] FTSE4Good Index,[75] Covalence EthicalQuote,[76] and AccountAbility's AA1000 standard[77] were created. Moreover, these actors are responsible for inquiries into the social networks' human rights compliance, as most prominently did 30 NGOs that issued a Letter to Social

[70] 'About GNI' (*Global Networks Initiative*) <https://globalnetworkinitiative.org/about-gni/> accessed 31 August 2022; Kathryn Doyle, 'BHR In The Tech Sector: Much To Celebrate, More To Do' <https://www.gp-digital.org/bhr-in-the-tech-sector-much-to-celebrate-more-to-do/> accessed 31 August 2022.

[71] 'Corporate Human Rights Policy' (*Meta*) <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf> accessed 31 August 2022; 'Our Commitment To Human Rights' (*Apple*) <https://s2.q4cdn.com/470004039/files/doc_downloads/gov_docs/Apple-Human-Rights-Policy.pdf> accessed 31 August 2022.

[72] 'Human Rights' (*Google*) <https://about.google/human-rights/> accessed 31 August 2022; 'Defending And Respecting The Rights Of People Using Our Service' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice> accessed 31 August 2022.

[73] 'Community Guidelines' (*TikTok*) <https://www.tiktok.com/community-guidelines#29> accessed 31 August 2022; 'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022.

[74] 'Creating A Common Good Balance Sheet' (*Economy for the Common Good*) <https://web.archive.org/web/20130426095936/http:/economia-del-bene-comune.it/en/content/creating-common-good-balance-sheet> accessed 31 August 2022.

[75] 'Index Inclusion Rules For The Ftse4good Index Series V2.0' (*FTSE Russell*) <https://web.archive.org/web/20171215102622/http:/www.ftse.com/products/downloads/F4G-Index-Inclusion-Rules.pdf> accessed 31 August 2022.

[76] 'About' (*Covalence*) <https://www.covalence.ch/index.php/about-us/> accessed 31 August 2022.

[77] 'FAQS' (*Corporate Register*) <https://www.corporateregister.com/about/> accessed 31 August 2022.

Media Platforms on Crisis Zones, which addressed the major social networks, used in the context of the Russian-Ukrainian War.[78] Despite the public attention to this attempt, only Twitter managed to produce a response.[79] Finally, to make a company socially responsible, users can engage in boycotting, ethical consumerism, and socially responsible investing to draw attention to the existing problems.[80] And if these approaches do not work out, there is always an option to bring the company to responsibility in a judicial manner, as was prominently done in *Milieudefensie et al v Royal Dutch Shell*.[81]

Overall, the implementation of UNGP's provisions remains a highly uneven sphere from the standpoints of states and the private sector alike. While Western democracies and a few countries from other regions actively introduce national action plans, the others remain ignorant of it. As for the businesses, the number of approaches is also quite vibrant even among the companies that belong to the same sector of social networks and are mainly incorporated in the US. Despite the active contributions from the social sector and users, UNGP adherence remains largely unregulated matter on the global level. However, as we will see in the forthcoming chapters, social networks tend to act more diligently to their human rights obligations in the countries with national action plans and specifically designed legislation. And to study this process in detail, we need to analyze the nuances of rules, tailored to the balancing

[78] 'Letter To Social Media Platforms On Crisis Zones' (*Electronic Frontier Foundation*, 2022) <https://www.eff.org/document/letter-social-media-platforms-crisis-zones> accessed 31 August 2022.

[79] 'NGOs Call On Social Media Platforms To Strengthen Human Rights Due Diligence In Crisis Situations; Incl. Co. Response' (*Business and Human Rights Resource Center*, 2022) <https://www.business-humanrights.org/en/latest-news/ngos-call-on-social-media-platforms-to-strengthen-their-human-rights-due-diligence-and-address-structural-inequalities-in-conflict-zones/> accessed 31 August 2022; 'Response From Twitter To Letter Calling On Social Media Platforms To Strengthen Their Human Rights Due Diligence' (*Business and Human Rights Resource Center*, 2022) <https://www.business-humanrights.org/en/latest-news/response-from-twitter-to-letter-calling-social-media-platforms-for-long-term-investment-in-human-rights/> accessed 31 August 2022.

[80] Markus Giesler and Ela Veresiu, 'Creating The Responsible Consumer: Moralistic Governance Regimes And Consumer Subjectivity' (2014) 41 Journal of Consumer Research; Andreas B. Eisingerich and others, 'Doing Good And Doing Better Despite Negative Information?: The Role Of Corporate Social Responsibility In Consumer Resistance To Negative Information' (2011) 14 Journal of Service Research; Bridget O'Laughlin, 'Governing Capital? Corporate Social Responsibility And The Limits Of Regulation' (2008) 39 Development and Change.

[81] NOS Nieuws, 'Milieudefensie Dagvaardt Shell In Rechtszaak Om Uitstoot' (2019) <https://nos.nl/artikel/2279155-milieudefensie-dagvaardt-shell-in-rechtszaak-om-uitstoot> accessed 31 August 2022; 'Climate Change Actions Against Corporations: Milieudefensie Et Al. V. Royal Dutch Shell Plc.' <https://www.cliffordchance.com/insights/resources/blogs/business-and-human-rights-insights/2021/01/climate-change-actions-against-corporations-milieudefensie-et-al-v-royal-dutch-shell-plc.html> accessed 31 August 2022.

of freedom of expression and incitement to violence and their ties to the content governance stakeholders.

II.B. Incitement to violence as an exception to freedom of expression

Freedom of expression is considered to be one of the pillars of democratic societies that gives people an opportunity to communicate their feelings to the masses, convey opinions on socio-political topics, criticize the actions of governments and contribute to the public discourse.[82] However, this freedom is not ultimate, as the power of words can be both a benefit and a hazard from the perspective of the values of communicated thoughts. The lessons of Nazi propaganda,[83] Rwandan radio transmissions,[84] and Russian hate-infused posts on social networks[85] give a clear outlook on the abuse of freedom of expression and urge the crystallization of perspicuous and sensible limitations on it. The operation of international organizations and courts greatly contributed to the emergence of balancing between freedom of expression and incitements to hate, violence, or even genocide. By outlining and understanding the principles of both freedom of expression and wrestling with incitements to violence, it is possible to identify the best balancing tests that would adhere to both. Consequently, in this section, I will highlight the major rules that govern the realms of freedom of expression and incitement to violence and determine the most comprehensive tests that should be applied in

[82] 'Freedom Of Expression - Article 10' (*Council of Europe*) <https://www.coe.int/en/web/human-rights-convention/expression> accessed 31 August 2022; Ashutosh Bhagwat and James Weinstein, 'Freedom Of Expression And Democracy', *The Oxford Handbook of Freedom of Speech* (Oxford Academic 2021); International IDEA, 'Press Freedom And The Global State Of Democracy Indices' (The Global State of Democracy in Focus 2019); Alicia Dibbets, Hans-Otto Sano and Marcel Zwamborn, 'Indicators In The Field Of Democracy And Human Rights: Mapping Of Existing Approaches And Proposals In View Of Sida's Policy' (The Danish Institute for Human Rights 2010); Marina Guseva and others, *Press Freedom And Development* (UNESCO 2008).

[83] Heidi Tworek, 'A Lesson From 1930S Germany: Beware State Control Of Social Media' *The Atlantic* (2019) <https://www.theatlantic.com/international/archive/2019/05/germany-war-radio-social-media/590149/> accessed 31 August 2022; The New Yorker, 'Copenhagen, Speech, And Violence' (2015) <https://www.newyorker.com/news/news-desk/copenhagen-speech-violence> accessed 31 August 2022; 'Nuremberg Trial Proceedings Vol. 12' (*The Avalon Project*) <https://avalon.law.yale.edu/imt/04-29-46.asp> accessed 31 August 2022; Wibke Kristin Timmermann, 'Incitement In International Criminal Law' (2006) 88 International Review of the Red Cross, p. 827.

[84] *Prosecutor v. Ruggiu* (Judgement and Sentence) ICTR-97-32-I (1 June 2000), paragraphs 16, 22; Wibke Kristin Timmermann, 'Incitement In International Criminal Law' (2006) 88 International Review of the Red Cross, p. 841.

[85] William Aceves, 'Virtual Hatred: How Russia Tried To Start A Race War In The United States' (2019) 24 Michigan Journal of Race & Law; Anne Applebaum, 'Ukraine And The Words That Lead To Mass Murder' *The Atlantic* (2022) <https://www.theatlantic.com/magazine/archive/2022/06/ukraine-mass-murder-hate-speech-soviet/629629/> accessed 31 August 2022.

the Internet domain by social networks to vitiate human rights violations and fulfill UNGP's obligations.

II.B.1. International standards for freedom of expression

The universally recognized concept of freedom of expression was proclaimed on an international level on 10 December 1948 with the adoption of the Universal Declaration of Human Rights, which stated that "everyone has the right to freedom of opinion and expression."[86] Another authoritative enlistment of this right was done in the International Covenant on Civil and Political Rights, where Article 19 includes a profound description of freedom of expression's scope and exercise.[87] This freedom was later reinstated in numerous international and regional documents confirming its universal importance in the human rights frameworks.[88] However, the ICCPR did not regulate the freedom of speech's application in the digital sphere, which gave rise to numerous operational and conceptual questions.

Concerning the essence of freedom of expression, it is a multifold concept, which includes both reception and sharing of information.[89] Consequently, this freedom can be viewed as a comprehensive concept that includes such rights as freedom of opinion, freedom of information, freedom of the press and the media, freedom of international communication, freedom of artistic expression, freedom of cultural expression, and freedom of science.[90] Importantly, the ways of information transmission do not matter, as the enforceability of this freedom is the same for oral, written, printed, or any other form of media.[91] Freedom of expression is also applicable to the sphere of the Internet, following the principle of "what

[86] Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR), Article 19.

[87] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR), Article 19.

[88] Council of Europe, European Convention on Human Rights (Rome, 4 November 1950) (ECHR), Article 10; Organization of African Unity, African (Banjul) Charter on Human and Peoples' Rights (27 June 1981) OAU Doc CAB/LEG/67/3, Article 9; Organization of American States, American Convention on Human Rights "Pact Of San Jose, Costa Rica" (B-32), Article 13; Association of Southeast Asian Nations, ASEAN Human Rights Declaration (19 November 2012), Principle 23.

[89] ICCPR, Article 19.

[90] Wolfgang Benedek and Matthias C. Kettemann, *Freedom Of Expression And The Internet* (Council of Europe Publishing 2013).

[91] ICCPR, Article 19.

applies offline also applies online."[92] Furthermore, this freedom is viewed as one of the safeguards of the Internet's openness and transparency, ensuring the free sharing of information between online actors.[93]

II.B.1.a. Recognition by the states

On the regional and national levels, freedom of expression achieved enormous recognition due to the majority of states' accession to the Universal Declaration of Human Rights.[94] Under the influence of the groundbreaking importance of freedom of expression, it is often cited as one of the fundamental rights in constitutions and other major legislative acts.[95] Moreover, numerous countries had laws dedicated to freedom of expression even before the adoption of UDHR. For instance, in the US it has been protected under the First Amendment to the United States Constitution since its adoption.[96] As of now, the freedom of expression is recognized by all European countries,[97] including the parties of the ongoing war, Ukraine[98] and Russia.[99] Thus, despite the question of the decency of this freedom's protection that regularly circulates in mass media and academia, its recognition on a state level constitutes a solid fact.

II.B.1.b. Apprehension by social networks

With regards to the major social networks, which are actively used during the course of the Russian-Ukrainian war, freedom of expression seems to play a prominent role in their day-to-day operations. For instance, Meta claims that it is committed to protecting the "voices" of

---

[92] Wolfgang Benedek and Matthias C. Kettemann, *Freedom Of Expression And The Internet* (Council of Europe Publishing 2013).

[93] 'Freedom Of Expression On The Internet' (*UNESCO*) <https://en.unesco.org/themes/freedom-expression-internet> accessed 31 August 2022.

[94] 'Human Rights Law' (*UN*) <https://www.un.org/ruleoflaw/thematic-areas/international-law-courts-tribunals/human-rights-law/> accessed 31 August 2022.

[95] Zachary Elkins and Tom Ginsburg, 'Imagining A World Without The Universal Declaration Of Human Rights' (2022) 74 World Politics, pp. 327-366; Hurst Hannum, 'The UDHR In National And International Law' (1998) 3 Health and Human Rights, pp. 151–152.

[96] U.S. Constitution, Amendment 1.

[97] Francesca Klug, *The Three Pillars Of Liberty: Political Rights And Freedoms In The United Kingdom* (Routledge 1996), p. 165; European Union, Charter of Fundamental Rights of the European Union (26 October 2012) 2012/C 326/02.

[98] Constitution of Ukraine, Article 34.

[99] Constitution of the Russian Federation, Article 29.

its users,[100] and Mark Zuckerberg, Meta's CEO, confirms to stand by "voice and free expression."[101] Moreover, free expression and sharing of people's content are mentioned as one of Facebook's and Instagram's main values in their community standards.[102] Twitter goes even further, mentioning major human rights documents, such as the Bill of Rights and the European Convention on Human Rights supplementing its commitment to freedom of expression.[103] A more interesting approach is in the case of Google, which mentions that all of its products are "for free expression",[104] however, no mentions of it can be found in YouTube community guidelines.[105] Telegram makes only two mentions of freedom of expression, stating that the content on the platform is not subject to local freedom of expression regulations and that the network's moderators "will not block anybody who peacefully expresses alternative opinions."[106] As to TikTok, its policies do not enlist freedom of expression among its values.[107] Consequently, the approaches to freedom of expression range from the elaborate commitments to freedom itself and the international documents, which proclaim it to the cases of total ignorance. As we will see in the next chapter, the manner in which platforms moderate their content is poles apart from the picture of freedom of expression recognition. Surprisingly, social networks with the strongest commitments to it seem to impose limitations most actively, contrary to social networks, which do not even mention it in their policies or mention it vaguely.

[100] 'We Are Committed To Protecting Your Voice And Helping You Connect And Share Safely' (*Meta,* 2022) <https://about.facebook.com/actions/promoting-safety-and-expression/> accessed 31 August 2022.

[101] 'Mark Zuckerberg Stands For Voice And Free Expression' (*Meta,* 2019) <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/> accessed 31 August 2022.

[102] 'Community Guidelines' (*Instagram*) <https://help.instagram.com/477434105621119/?helpref=hc_fnav> accessed 31 August 2022; 'Facebook Community Standards' (*Facebook*) <https://transparency.fb.com/policies/community-standards/> accessed 31 August 2022.

[103] 'Defending And Respecting The Rights Of People Using Our Service' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice> accessed 31 August 2022.

[104] 'Terms And Policies' (*Google*) <https://support.google.com/googlecurrents/answer/9680387?hl=en> accessed 31 August 2022.

[105] 'Community Guidelines' (*YouTube*) <https://www.youtube.com/intl/en_us/howyoutubeworks/policies/community-guidelines/#community-guidelines> accessed 31 August 2022.

[106] 'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022.

[107] 'Community Guidelines' (*TikTok*) <https://www.tiktok.com/community-guidelines#29> accessed 31 August 2022.

II.B.2. Elements of incitement to violence

Hate speech is widely recognized as one of the major hazards to human rights and the rule of law.[108] Despite the grave nature of this form of speech, the global community still lacks a universally accepted definition.[109] The major reason for this is the broad concept of hatred, which can have various meanings depending on culture, religion, language, and other local specificities.[110] One of the documents that tried to define the specific forms of hate speech is the Additional Protocol to the Convention on Cybercrime, which prohibits "racist and xenophobic material."[111] As for the definition itself, this category includes "any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, color, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors."[112] Another take on hate speech was accomplished by the UN in its 2019 Strategy and Plan of Action on Hate Speech, where it is defined as "any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender, or other identity factor."[113] Despite the international nature of such documents, the eagerness of countries to limit the effects of hate speech has remained scarce, with few exceptions such as The EU Code of conduct on countering illegal hate speech online. This document obliged social networks to monitor and take action regarding hate speech.[114] However, even this document does not have

[108] 'Hate Speech Is Rising Around The World' (*UN*) <https://www.un.org/en/hate-speech> accessed 31 August 2022.

[109] UN Committee on the Elimination of Racial Discrimination, 'General Recommendation No. 32 on The Meaning and Scope of Special Measures in the International Convention on the Elimination of Racial Discrimination' (24 September 2009) UN Doc CERD/C/GC/32, paragraph 9; Natalie Alkiviadou, 'Hate Speech On Social Media Networks: Towards A Regulatory Framework?' (2019) 28 Information & Communications Technology Law, p. 22.

[110] Ibid.

[111] Council of Europe, Additional Protocol to the Convention on Cybercrime (Strasbourg, 28 January 2003) European Treaty Series - No. 189, Article 1.

[112] Ibid.

[113] 'United Nations Strategy And Plan Of Action On Hate Speech' (*UN*, 2019) <https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf> accessed 31 August 2022.

[114] 'The EU Code Of Conduct On Countering Illegal Hate Speech Online' (*European Commission*) <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-

strict enforcement mechanisms, which undermines the efficacy of this instrument. As it is visible from the latest report, not all the users' reports were assessed in less than 24 hours and even the content with messages about killing and violence was deleted only in 69% of cases.[115]

Perhaps, one of the reasons for such unwillingness of stakeholders to combat hate speech is the broad nature of this category, the assessment of which entails a great degree of subjectivity. One of the ways to limit the negative impact of hate speech and maximize the speed and objectivity of reaction may be the application of internationally recognized standards with a steady practice of use. Such examples may be found in the category of incitement to violence and, particularly, genocide. Both of these types of incitement are parts of a greater hate speech concept.[116] However, their inchoate nature led to their broad recognition in both natural and international legislation.[117]

Regarding the major definitions of this crime, according to the UN Human Rights Office of the High Commissioner (further – "OHCHR"), incitement to hatred has three important terms: 1) "hatred" and "hostility" ("intense and irrational emotions of opprobrium, enmity and detestation towards the target group"), 2) "advocacy" ("requiring an intention to promote hatred publicly towards the target group") and 3) "incitement" ("statements about national, racial or religious groups, which create an imminent risk of discrimination, hostility or violence against persons belonging to those groups").[118] As for the elements of the crime, incitement to genocide has the most authoritative definitions, as they were given by international criminal tribunals. As for objective elements, there should be an actual incitement to genocide of direct and public nature, targeted at a specific group.[119] Concerning *mens rea*, it entails the intent to incite others

xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct> accessed 31 August 2022.

[115] Didier Reynders, 'Countering Illegal Hate Speech Online: 6Th Evaluation Of The Code Of Conduct' (European Commission 2021), p. 2.

[116] Angela Hefti and Laura Ausserladscheider Jonas, 'From Hate Speech To Incitement To Genocide: The Role Of The Media In The Rwandan Genocide' (2020) 38 Boston University International Law Journal.

[117] UNGA Res 60/1 (24 October 2005) UN Doc A/RES/60/1, paragraphs 138-139; UNGA 'Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred' (11 January 2013) UN Doc A/HRC/22/17/Add.4; 18 U.S. Code § 1091 – Genocide; UNGA Rome Statute of the International Criminal Court (adopted 17 July 1998, last amended 2010) UN Treaty Series vol. 2187 no. 38544 (Rome Statute), Article 25(3)(e).

[118] UNGA 'Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred' (11 January 2013) UN Doc A/HRC/22/17/Add.4.

[119] UNSC Statute of the International Criminal Tribunal for Rwanda (adopted 8 November 1994 UNSC Res S/RES/955, last amended 2006) (ICTR), Article 2(3)(c); UNSC Statute of the International Criminal Tribunal for

to commit genocide.[120] The case of *Ruggiu* is one of the most prominent examples that satisfied this test and can serve as a benchmark for the situations in this category. As per the facts of this case, Ruggiu was the employee of RLTM radio, the influential and popular mass media platform.[121] Consequently, Ruggiu transmitted messages to destroy and kill Tutsis, which eventually resulted in the actual genocide of this group.

Nowadays, the power of the Internet is much greater than that of local radio in 1994,[122] and it is regarded as public media, where numerous atrocities regularly occur.[123] Therefore, there are no limitations to introducing the aforementioned tests to the sphere of the Web, as they will provide a detailed outline of monitoring for the moderators. However, modern social networks adhere to different standards and regularly construct their own definitions of hate speech. And, as we will see in the next subsection, the balancing tests between incitement to violence and freedom of expression in the digital realm serve as a compromise between the vague terminology of hate speech and the massive authority of freedom of expression.

II.B.2.a. Incitement to violence on the level of national legislation

On the state level, incitement to violence almost always constitutes a criminal offense, although the naming of the crime and its elements may be slightly different. For instance, in the EU only two countries do not have specific legislative provisions against any form of hate speech.[124] In the meantime, such countries as Germany, France, Spain, Sweden, and many others, have outlawed incitement to violence on a variety of grounds, including sex, nationality, race, and other prerequisites.[125] Moreover, incitement to violence is prohibited in common law

---

the Former Yugoslavia (adopted 25 May 1993 UNSC Res 827/1993, last amended 17 May 2002) (ICTY), Article 4(3)(c); *Nahimana et al. (Media case)* (Appeal Judgement) ICTR-99-52-A (28 November 2007), paragraph 677; *Ngirabatware Augustin* (Appeal Judgement) MICT-12-29-A (18 December 2014), paragraph 52.

[120] Ibid.

[121] *Prosecutor v. Ruggiu* (Judgement and Sentence) ICTR-97-32-I (1 June 2000), paragraphs 17-44.

[122] Digital Vs Traditional Media Consumption' (*globalwebindex*) <https://www.amic.media/media/files/file_352_2142.pdf> accessed 31 August 2022.

[123] Federal Bureau of Investigation, 'Internet Crime Report 2021' (Federal Bureau of Investigation 2021).

[124] ERR, 'Estonia One Of Two EU Countries Not To Criminalize Hate Speech' (2020) <https://news.err.ee/1159938/estonia-one-of-two-eu-countries-not-to-criminalize-hate-speech> accessed 31 August 2022.

[125] Strafgesetzbuch 1975, §130 (1); Loi 90-615 du 13 juillet 1990; Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, 510; Brottsbalk (1962:700), 16(8).

countries, including the US and the UK.[126] As to the parties for the parties of the war, incitement is also criminalized in their legislation. For instance, in Russia "actions aimed at the incitement of hatred or enmity" are prohibited under Article 282 of the Criminal Code.[127] Concerning Ukraine, protection against hate crimes is granted on both constitutional and criminal levels, outlawing both the action of incitement and the manufacture and distribution of materials that promote it.[128] Consequently, incitement to violence is almost universally condemned with minor differences in its wording. However, the implementation of criminal norms may greatly vary due to the differences in political regimes and the deployed tests that balance inciting speech against freedom of expression.

II.B.2.b. The social networks' perspective on incitement to violence

Incitement to violence and hate speech are represented in the social networks' policies even more prominently than the freedom of expression. Starting from Meta, it mentions the types of content, which are not allowed right after the passage concerning the importance of free speech.[129] Facebook Community Standards address such topics as "Violence and Incitement" and "Hate Speech", explaining them as "language that incites or facilitates serious violence"[130] and "a direct attack against people"[131] respectively. Moreover, on these pages, users are able to find the policy rationale for the prohibition of such content, useful examples, reports, and instruments, available to users to fight illegal content on the platform's premises. As to the other product of Meta, Instagram, its policy directly states that "Instagram is not a place to support or praise terrorism, organized crime, or hate groups."[132] Interestingly, by

[126] Thomas R Hensley, *The Boundaries Of Freedom Of Expression & Order In American Democracy* (Kent State University Press 2001), p. 153; Public Order Act 1986; Criminal Justice Act 2003; Racial and Religious Hatred Act 2006 (England and Wales); 18 U.S. Code § 373 - Solicitation to commit a crime of violence.

[127] State Duma, Criminal Code of the Russian Federation, Article 282.

[128] Constitution of Ukraine, Articles 24, 37; Verkhovna Rada of Ukraine, Criminal Code of Ukraine, Articles 161, 300.

[129] 'Facebook Community Standards' (*Facebook*) <https://transparency.fb.com/policies/community-standards/> accessed 31 August 2022.

[130] 'Violence And Incitement' (*Meta*) <https://transparency.fb.com/policies/community-standards/violence-incitement/#policy-details> accessed 31 August 2022.

[131] 'Hate Speech' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/> accessed 31 August 2022.

[132] 'Community Guidelines' (*Instagram*) <https://help.instagram.com/477434105621119/?helpref=hc_fnav> accessed 31 August 2022.

clicking on such phrases as “hate speech” or “serious threats or harm”, the user is sent to the aforementioned “Hate Speech”[133] and “Violence and Incitement”[134] pages on Facebook Community guidelines, meaning that their definitions are the same in Meta, with no diversification as to the products. Instagram also mentions the only instrument, available to users concerning combating violations of their standards under the name of “Legal Removal Request.”[135] As to the statistics on content moderation, they can be found in the same report, provided on Facebook’s “Hate Speech” page.[136]

Moving to Twitter, it establishes violence as its primary concern about safety on its platform. The platform mentions that such actions as “Incitement against protected categories” is prohibited by the tandem of “Violent threat” and “Hateful conduct” policy, and is defined by Twitter as an “inciting behavior that targets individuals or groups of people belonging to protected categories.”[137] The policy offer examples to diversify between legal and infringing content and provide steps for user reporting of violations. However, the platform does not provide statistics on its enforcement mechanisms.

With respect to YouTube, it provides a “Hate speech” policy, where hate speech is defined as “content promoting violence or hatred against individuals or groups” further enlisting the protected groups.[138] Moreover, YouTube offers a video on this topic, which represents their view on the problem.[139] Examples, user-reporting mechanisms, and enforcement statistics are also provided on the page.

---

[133] 'Hate Speech' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/> accessed 31 August 2022.

[134] 'Violence And Incitement' (*Meta*) <https://transparency.fb.com/policies/community-standards/violence-incitement/#policy-details> accessed 31 August 2022.

[135] 'Legal Removal Request' *(Instagram)* <https://help.instagram.com/874680996209917/?helpref=hc_fnav> accessed 31 August 2022.

[136] 'Hate Speech – Data' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/#data> accessed 31 August 2022.

[137] 'Hateful Conduct Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy> accessed 31 August 2022.

[138] 'Hate Speech Policy' (YouTube Help) <https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436> accessed 31 August 2022.

[139] Ibid.

TikTok enlists two types of prohibited content, which can be linked to incitement to violence: "Hateful behavior"[140] and "Violent extremism."[141] The first category includes hate speech and is explained as "content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group" with the respective list of the groups.[142] "Violent extremism" covers incitement to violence and means "advocating for, directing, or encouraging other people to commit violence."[143] Also, TikTok mentions the types of content, which are covered by these policies. Apart from that, no information is given to the users.

As to Telegram, it mentions only a vague category of illegal content, which can be reported by users.[144] As for the definition of such category, enforcement actions, and inclusion of incitement to violence or hate speech categories, no details are given to the users.[145]

II.B.3. A walk on thin ice: Balancing freedom of expression and incitement to violence

Freedom of speech and prohibition of incitement to violence are two universally recognized principles, which are yet to find a proper balancing mechanism between each other. The difficulties may stem from the absence of a unified document, dedicated to this issue, or from the differences in various national and international regimes that try to weigh these two notions. However, even with some complications, it is possible to observe a couple of predominant tendencies, inherent to this sphere: 1) freedom of expression is usually regarded as a paramount right, 2) incitement to violence is a unilaterally prohibited action and is viewed as an exception to freedom of expression, 3) to ban a certain inciting message it is obligatory to balance it with freedom of expression. This view is supported by a number of international and national documents, judicial decisions, and statements of major legal and political institutions. Acknowledging these facts, the focus will be put on the aforementioned sources to find out the most applicable tests, which help to decide whether a certain massage is inciting and should be dealt with.

[140] 'Hateful Behavior' (*TikTok*) <https://www.tiktok.com/community-guidelines#38> accessed 31 August 2022.

[141] 'Violent Extremism' (*TikTok*) <https://www.tiktok.com/community-guidelines#39> accessed 31 August 2022.

[142] 'Hateful Behavior' (*TikTok*) <https://www.tiktok.com/community-guidelines#38> accessed 31 August 2022.

[143] 'Violent Extremism' (*TikTok*) <https://www.tiktok.com/community-guidelines#39> accessed 31 August 2022.

[144] 'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022.

[145] Ibid.

II.B.3.a. The international dimension

Starting from the international document of the utmost importance, ICCPR contains two notions that relate to limitations of freedom of speech (Articles 19 and 20). The provision with a direct link to incitement to violence is Article 20(2), which states that the "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."[146] Despite the clear wording of these articles, it is rarely enlisted in the same manner in regional and national legislation, often opting for weaker or simply different wording.[147] In this regard, the courts act as primary identifiers of the relevant rules and are the most efficient actors in the enforcement of written provisions.[148] The standard is regularly applied in conjunction with Article 19 ICCPR, forming a specific test, which takes into consideration the following prerequisites: 1) legality (whether a limitation is provided by law),[149] 2) legitimacy (whether a limitation pursues a legitimate aim),[150] and 3) necessity (whether a limitation is necessary in a democratic society).[151]

Another element, which is not directly addressed in ICCPR, but is incredibly relevant for the judicial assessment is the threshold test. Through its application, it is possible to diversify the speech and messages that should be assessed through the prism of incitement to violence and others that do not reach the same level of gravity and viciousness.[152] To help the courts accomplish such a task, OHCHR came up with the Rabat Plan of Action, which

[146] ICCPR, Article 20.

[147] Barbora Bukovska, Agnes Callamard and Sejal Parrmer, 'Towards An Interpretation Of Article 20 Of The ICCPR: Thresholds For The Prohibition Of Incitement To Hatred' (Article 19 2010) <https://www2.ohchr.org/english/issues/opinion/articles1920_iccpr/docs/CRP7Callamard.pdf> accessed 31 August 2022, pp. 3-5.

[148] Ibid., p. 18.

[149] *The Sunday Times v. United Kingdom* App no 6538/74 (ECtHR 26 April 1979), paragraph 49.

[150] Barbora Bukovska, Agnes Callamard and Sejal Parrmer, 'Towards An Interpretation Of Article 20 Of The ICCPR: Thresholds For The Prohibition Of Incitement To Hatred' (Article 19 2010) <https://www2.ohchr.org/english/issues/opinion/articles1920_iccpr/docs/CRP7Callamard.pdf> accessed 31 August 2022, pp. 3-5.

[151] *Zana v Turkey* App no 18954/91 (ECtHR 25 November 1997), paragraph 51; *Lingens v Austria* App no 9815/82 (ECtHR 8 July 1986), paragraphs 39-40.

[152] Asma Jahangir and Doudou Diène, 'Report Of The Special Rapporteur On Freedom Of Religion Or Belief, Asma Jahangir, And The Special Rapporteur On Contemporary Forms Of Racism, Racial Discrimination, Xenophobia And Related Intolerance, Doudou Diène, Further To Human Rights Council Decision 1/107 On Incitement To Racial And Religious Hatred And The Promotion Of Tolerance' (UN Human Rights Council 2022) <https://digitallibrary.un.org/record/583355> accessed 31 August 2022.

specifically indicates seven criteria that should be assessed to determine whether a certain statement amount to a criminal offense. Namely, the document enlists the following elements: 1) context, 2) speaker, 3) intent, 4) content and form, 5) the extent of the speech act, and 6) likelihood, including imminence.[153]

The aforementioned positions are regularly used by the European Court of Human Rights and are considered to be the highest standards, set in this sphere. However, national legal systems can adopt vastly different interpretations or completely disregard these notions with no repercussions from OHCHR. Partly it can be explained by the non-obligatory nature of the Rabat Plan of Action and the broad nature of Article 20(2) ICCPR. However, the primary role would probably be played by national legal traditions, the political interests of the governing parties, and current political regimes. Eventually, it would be virtually impossible to create a unified set of obligations that would be identically applicable to the whole online sphere, including the biggest social networks.

II.B.3.b. The EU perspective

The national legal dimension of the balancing between freedom of expression and incitement to violence differs from country to country, which is explained by such factors as differences in legal systems, roles of courts and legislators, and many other legal pretexts. For instance, in the US the balancing functions have been usually performed by the courts, which led to the development of such tests as clear and present danger, bad tendency, and, ultimately, incitement to imminent lawless action.[154] However, as we have already seen in chapter II.B.2.a., even common law countries perceive incitement to violence as an inchoate crime and criminalize it in national legislation. Therefore, these provisions by default help to bypass freedom of expression considerations if certain speech satisfies all the elements of the crime, making it an automatic exclusion from freedom of expression protection. It is worth mentioning that the majority of countries do not have legal norms, specifically tailored to combating incitement to violence in the digital sphere. The only countries, which have successfully

[153] UNGA 'Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred' (11 January 2013) UN Doc A/HRC/22/17/Add.4.

[154] John R. Vile, 'Incitement To Imminent Lawless Action' <https://www.mtsu.edu/first-amendment/article/970/incitement-to-imminent-lawless-action> accessed 31 August 2022.

adopted online-focused legislation are Germany and Austria, which came up, respectively with NetzDG and CPA.

The aforementioned acts share several distinct features, which ensure effective regulation and must be adhered to by the online platforms that operate in Germany and Austria. First of all, they deploy the strategy of intermediary liability, putting the responsibility over safety on the Internet on social networks and video sharing platforms. As was established in section I.A.1., the acts define the scope of actors, responsible for the fulfillment of monitoring and acting obligations, diversifying between regular communication platforms (social networks) and video sharing platforms. Both these groups should conduct two sets of actions: react to complaints regarding illegal content and report to the authorities on the actions taken to combat such content.[155] As to the nature of illegal content, both acts stick to the definitions in national legislation, providing links to the criminal codes of their countries. Incitement to violence is penalized in NetzDG and CPA, being formulated as "incitement of masses" under Section 130 of German StGB[156] and "incitement" under § 283 of Austrian StGB.[157] Both acts underline that the platforms must have clear and easily understandable mechanisms for user reporting.[158] After receiving the reports platforms have to review and deleted the illegal content within 24 hours or seven days if the review process requires additional analysis.[159] The review procedures must also be transparent and effective. NetzDG even comes up with a requirement of regular German language training for people, who conduct the review procedures.[160] Finally, under NetzDG and CPA, the actions of relevant platforms are subject to oversight from national authorities and fines can be imposed in cases of non-compliance.[161]

For now, NetzDG and CPA remain the only European acts that explicitly regulate the monitoring activities of social networks. However, such law as DSA is currently on the horizon of EU legislation and will soon extrapolate the monitoring obligations to the level of EU

[155] CPA, § 4; NetzDG, § 2.

[156] NetzDG, § 1(3).

[157] CPA, § 2(8).

[158] Ibid., § 4; NetzDG, § 2.

[159] Ibid.

[160] NetzDG, § 3(5).

[161] Ibid., § 4a; CPA, § 8.

Member States.[162] Similarly to German and Austrian laws, this document also contains the definitions of relevant actors,[163] illegal content,[164] information about monitoring and reporting obligations,[165] fines,[166] and governing authority.[167] However, in all of these categories, it tends to broaden the regulatory scope and adds a few interesting details. For instance, the DSA contains the broad umbrella term of "intermediaries", which can perform "mere conduit", "caching" and "hosting" functions.[168] The social networks, used in the context of the Russian-Ukrainian war belong to the "hosting" category, which includes online platforms. As to the latter category, the DSA defines it as "a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information, unless that activity is a minor and purely ancillary feature of another service and, for objective and technical reasons cannot be used without that other service, and the integration of the feature into the other service is not a means to circumvent the applicability of this Regulation."[169] Moreover, online platforms contain one more sub-category of "very large online platforms", which is linked to the number of users from EU.[170] The obligations are imposed on the actors according to their belonging to one of the aforementioned groups. In this regard, we can outline the most important obligations of the regular online platform:

- It should act against illegal content and provide necessary information on receipt of the order from national authorities;[171]

[162] 'Digital Services: Landmark Rules Adopted For A Safer, Open Online Environment' (*European Parliament*, 2022) <https://www.europarl.europa.eu/news/en/press-room/20220701IPR34364/digital-services-landmark-rules-adopted-for-a-safer-open-online-environment> accessed 31 August 2022.

[163] DSA, Article 2.

[164] Ibid., Article 2(g).

[165] Ibid., Article 13, 14.

[166] Ibid., Article 42.

[167] Ibid., Article 38.

[168] Ibid., Article 2(f).

[169] Ibid., Article 2(h).

[170] Ibid., Article 25.

[171] Ibid., Articles 8, 9.

- Must define all restrictions on the content and users' actions in the terms and conditions;[172]
- Is obliged to have notice and action mechanisms to combat illegal content;[173]
- Shall provide an "easy to access, user-friendly" complaint-handling system to address the reports of their users;[174]
- Should issue reports as to their activities against illegal content.[175]

Moreover, the category of very big online platforms should also conduct:

- Risk assessment activities (evaluate their current content moderation activities);[176]
- Risk mitigation (ensure that their mechanisms are in line with combating illegal content);[177]
- Audits (independent checks as to the efficiency of their procedure);[178]
- Employment of at least one compliance officer to oversee conformity with legal standards.[179]

In this regard, the DSA draws some inspiration from UNGP, which enlists risk assessment as one of the main mechanisms for business compliance with human rights standards.

As to the definition of illegal content, the DSA does not go into the details, simply stating that it is any information that "is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law."[180] Consequently,

[172] Ibid., Article 12.

[173] Ibid., Article 14.

[174] Ibid., Article 17.

[175] Ibid., Article 23.

[176] Ibid., Article 26.

[177] Ibid., Article 27.

[178] Ibid., Article 28.

[179] Ibid., Article 32.

[180] Ibid., Article 2(g).

incitement to violence should be also in the scope of infringing content, as it can be found in legislative acts of the majority of EU members.

The DSA contains two more interesting details, which are absent from NetzDG and CPA. In particular, it reiterates the importance of freedom of expression and emphasizes that all activities should be done with this universal freedom in mind.[181] However, the Act does not stipulate specific balancing tests, limiting itself by simple consideration of freedom. Another element is the mention of the importance of multistakeholder cooperation to achieve the goal of the DSA. In this regard, the drafters mention the consultations with online platforms, the civil sector, and online users, whose opinions were taken into consideration during the drafting process of the DSA.[182] The drafters also highlight the utmost importance of multistakeholder cooperation for the successful enforcement of this Act.

It should be noted that NetzDG, CPA, and DSA have all received a fair portion of criticism, including the threats to freedom of speech, creation of impediments to business operations, unnecessity pursuant to the prior existence of reporting mechanisms and the absence of substantial additions to the existing mode of action towards illegal online content.[183] Such claims have their actual prerequisites and are truly important for the efficient modernization of legislative framework in this field. Moreover, such critique is a clear example of multistakeholder approach, in which different actors participate in the framing of a better answer to the itchy topic. However, all the drawbacks of these legislative acts do not undermine the importance of their analysis for the sake of combating incitement to violence. Hopefully, legislators will address the problematic points, which have already been highlighted, and will design laws with greater efficiency. But for now, these laws remain the only European acts that

[181] Ibid., Article 26(1)(b); Explanatory Memorandum to DSA, Sections 2, 3.

[182] Explanatory Memorandum to DSA, Section 3.

[183] Amélie Heldt, 'Reading Between The Lines And The Numbers: An Analysis Of The First Netzdg Reports' (2019) 8 Internet Policy Review; Gabriela Staber, 'Communication Platforms Face New Obligations And High Fines In Austria' <https://www.lexology.com/library/detail.aspx?g=fcf46df4-4694-4f10-b11b-67564a824470> accessed 31 August 2022; 'Austria: The Draft Communication Platforms' Act Fails To Protect Freedom Of Expression' <https://www.article19.org/resources/austria-draft-communication-platforms-act-fails-freedom-of-expression/> accessed 31 August 2022; Shane O'Callaghan, 'Can The EU Digital Services Act Succeed In Controlling Big Tech And Protecting Consumer Rights?' <https://www.iglobenews.org/can-the-eu-digital-services-act-succeed-in-controlling-big-tech-and-protecting-consumer-rights/> accessed 31 August 2022; Ilaria Buri and Joris van Hoboken, 'The Digital Services Act (DSA) Proposal: A Critical Overview' (Digital Services Act (DSA) Observatory, Institute for Information Law (IViR), University of Amsterdam 2021) <https://dsa-observatory.eu/wp-content/uploads/2021/11/Buri-Van-Hoboken-DSA-discussion-paper-Version-28_10_21.pdf> accessed 31 August 2022.

somehow address the existing issue and modernize their legislative framework so that it would correspond to digital standards. Even if their approaches are not free from imperfections, they still serve as an important initial step and should be scrutinized by the representatives of various disciplines.

II.B.3.c. Approaches of states at war

As to the actions of Russia and Ukraine, both states have taken steps that concerned the question of incitement to violence and freedom of expression, but these actions were highly influenced by the governmental perception of good and evil. Russian legislative acts have been an example of condemnation and prohibition of independent Western and local media, which highlighted the criminal actions of Russian officials and the army.[184] For example, Russian legislation criminalized such actions, as the spread of "unreliable information" concerning state officials and armed forces,[185] calls for sanctions against Russia, and the dissemination of "extremist" information and "fake news."[186] The results of these laws' enforcement include the imprisonment of members of the opposition, a ban on independent news sources, and the withdrawal of human rights NGOs and media platforms from the country.[187] As for Ukraine, it imposed the ban on Russian social networks, which prominently hosted the calls for violence against Ukrainians and were powerful means of Russian anti-democratic propaganda.[188] Furthermore, Ukrainian state officials made several inquiries and public pledges to such online

[184] 'Censorship: Russia Blocks Access To Independent Media Over War Coverage' <https://europeanjournalists.org/blog/2022/03/01/censorhip-russia-blocks-access-to-independent-media-over-war-coverage/> accessed 31 August 2022; 'Russia's Crackdown On Independent Media And Access To Information Online' <https://www.csis.org/analysis/russias-crackdown-independent-media-and-access-information-online> accessed 31 August 2022; Pjotr Sauer, 'Russia Bans Facebook And Instagram Under 'Extremism' Law' *The Guardian* (2022) <https://www.theguardian.com/world/2022/mar/21/russia-bans-facebook-and-instagram-under-extremism-law> accessed 31 August 2022.

[185] State Duma, Federal Law No. 31-FZ of 4 March 2022.

[186] Ibid.; 'Russia Criminalizes Independent War Reporting, Anti-War Protests' <https://www.hrw.org/news/2022/03/07/russia-criminalizes-independent-war-reporting-anti-war-protests> accessed 31 August 2022.

[187] Radio Liberty, 'Putin Signs 'Harsh' Law Allowing Long Prison Terms For 'False News' About Army' (2022) <https://www.rferl.org/a/russia-military-false-news/31737627.html> accessed 31 August 2022; Michael M. Grynbaum, John Koblin and Tiffany Hsu, 'Several Western News Organizations Suspend Operations In Russia' *The New York Times* (2022) <https://www.nytimes.com/2022/03/04/business/western-media-operations-russia.html> accessed 31 August 2022; Chloe Folmar, 'Multiple NGOs Including Amnesty International Forced To Shutter Offices In Russia' *The Hill* (2022) <https://thehill.com/policy/international/russia/3263143-multiple-ngos-including-amnesty-international-forced-to-shutter-offices-in-russia/> accessed 31 August 2022.

[188] Alec Luhn, 'Ukraine blocks popular social networks as part of sanctions on Russia' *The Guardian* (2017).

intermediaries as Meta and Twitter to adjust their moderation activities to protect the safety of Ukrainian users and delete the content, which infringes human rights.[189]

The response from users and the social sector has been also highly uneven, with some voices proclaiming the importance of legislation against online hate speech,[190] and others arguing about an ultimate danger to freedom of expression such laws create.[191] However, in the context of the Russian-Ukrainian war, the agenda is slowly shifting towards consensus between the social sector and some users concerning the ban on incitement to violence against Ukrainians, which has been supported by numerous reports, studies, online campaigns, and public outcries against social platforms' inaction.[192]

Consequently, in the majority of countries, the legislative framework concerning balancing freedom of expression and incitement to violence is mostly limited to general norms, which do not provide accountability for online platforms. However, some countries managed to adopt exclusive acts, which enforce the concept of intermediary liability, such as NetzDG and CPA. Despite being limited by the jurisdictions of the two countries, taking into account the issues of drawing the lines between national jurisdictions in the online sphere and a big number of users from Germany and Austria, online platforms tend to adapt their standards uniformly, instead of looking for a country-specific approach. Moreover, social networks have

[189] Will Oremus, 'Ukraine Says Big Tech Has Dropped The Ball On Russian Propaganda' *The Washington Post* (2022) <https://www.washingtonpost.com/technology/2022/07/14/ukraine-takedown-requests-russia-propaganda/> accessed 31 August 2022; Interfax-Ukraine, 'Ukraine's Dpty PM Asks Intl Services, Social Networks To Block Their Content In Russia' (2022) <https://en.interfax.com.ua/news/economic/802555.html> accessed 31 August 2022.

[190] Susan Benesch and others, *Reducing Online Hate Speech: Recommendations For Social Media Companies And Internet Intermediaries* (The Israel Democracy Institute 2020).

[191] Article 19, 'Responding To 'Hate Speech': Comparative Overview Of Six EU Countries' (Article 19 2018) <https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf> accessed 31 August 2022.

[192] 'Updates: Digital Rights In The Russia-Ukraine Conflict' <https://www.accessnow.org/digital-rights-ukraine-russia-conflict/> accessed 31 August 2022; 'Ukraine: Briefing On 'Incitement To Violence Leading To Atrocity Crimes' <https://reliefweb.int/report/ukraine/ukraine-briefing-incitement-violence-leading-atrocity-crimes> accessed 31 August 2022; Ukrinform, 'Russia's State-Orchestrated Incitement To Genocide Of Ukrainians' (2022) <https://www.ukrinform.net/rubric-ato/3495013-russias-stateorchestrated-incitement-to-genocide-of-ukrainians.html> accessed 31 August 2022; Republicworld, 'UNSC To Consider 'Incitement To Violence' Among Reasons Behind Russia's War In Ukraine' (2022) <https://www.republicworld.com/world-news/russia-ukraine-crisis/unsc-to-consider-incitement-to-violence-among-reasons-behind-russias-war-in-ukraine-articleshow.html> accessed 31 August 2022; Ivana Kottasova, 'Russia Accused Of Inciting Genocide In Ukraine In New Report' *CNN* (2022) <https://www.ctvnews.ca/world/russia-accused-of-inciting-genocide-in-ukraine-in-new-report-1.5920894> accessed 31 August 2022; Claire Parker, 'Russia Has Incited Genocide In Ukraine, Independent Experts Conclude' *The Washington Post* (2022) <https://www.washingtonpost.com/world/2022/05/27/genocide-ukraine-russia-analysis/> accessed 31 August 2022.

to acknowledge the importance of such forthcoming legislative novelty as DSA, which will make compliance obligations more stringent and harder to vitiate from. Also, there is pressure from the public sector and users who urge online companies to introduce more stringent responses and adapt their moderation activities to the current needs. With this pretext in mind, we can comprehensively analyze the current moderation activities of social networks against incitement to violence during the Russian-Ukrainian war and the influence of internet governance actors on them.

III. THE LESSONS FROM THE RUSSIAN-UKRAINIAN WAR: ASSESSING SOCIAL NETWORKS' MODERATION ACTIVITIES

As we have already found out in the previous chapters, all the actors in the system of multistakeholder digital content governance, recognize and implement such constructs, as an obligation to protect human rights, freedom of expression, and prohibition of incitement to violence. Naturally, such recognition highly differs in the levels of legal power, enforceability, and contribution to the current state of affairs. As the current Russian-Ukrainian war has demonstrated, one of the most problematic topics is finding a balance between respect for freedom of expression and prohibition of incitement to violence. In addition to the absence of obligatory legislation in the majority of countries, there is incoherence in the opinions of public sector actors and users with slightly ambivalent appreciations of freedom of expression and incitement to violence. In this scenario, it would be hard for social networks to introduce decent reactive mechanisms to infringing content. However, the legislators of Germany, Austria, and the EU managed to come up with regulations for the combating of digital incitement to violence with such laws as NetzDG, CPA, and DSA. Even prior to that such social networks as Facebook, Instagram, YouTube, Twitter, and TikTok expressed their commitment to combat hate speech in the agreement with the European Commission under the name of the Code of conduct on countering illegal hate speech online.[193] Despite the arising proactive legislative framework, the implementation of the aforementioned norms by social networks remains highly uneven both on declarative and operational levels. Therefore, to properly assess the levels of social networks' compliance we will focus on six networks, which play the most crucial role in the current Russian-Ukrainian war: Facebook, Instagram, Twitter, YouTube, TikTok, and Telegram. The points for analysis would be 1) the definitions of tests for balancing freedom of expression and incitement to violence in the policies, 2) issuing of reports on combating illegal content, 3) actions taken in the course of the Russian-Ukrainian war, 4) availability of users notice functionality, 5) presence of infringing content and 6) reaction to users' reports on the content with presumed incitement to violence.

[193] 'The EU Code Of Conduct On Countering Illegal Hate Speech Online' (*European Commission*) <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct> accessed 31 August 2022.

III.A. Social networks' activities

III.A.1 Meta

III.A.1.a. Policy definitions and reports

Analyzing the definitions and representation of both freedom of expression and incitement to violence, Meta would be a social network with a high degree of clarity as to the actions that fall into the scope of prohibited content. The definitions and guidelines for users are the same for Facebook and Instagram since both social networks are administered by the same company – Meta.

As to the scope of incitement to violence, it is represented by such Meta policies, as "Hate Speech"[194] and "Violence and Incitement."[195] Both of these policies contain examples that are automatically banned with no freedom of expression consideration due to their hazardous nature and the ones that are assessed in relation to the context and additional information since they can express political views and constitute a matter of public interest. Automatically banned hate speech is defined as "content targeting a person or group of people (including all groups except those who are considered non-protected groups described as having carried out violent crimes or sexual offenses or representing less than half of a group) on the basis of their aforementioned protected characteristic(s) or immigration status" performed by such actions as violent or dehumanizing speech, mocking of the victims, dehumanizing comparisons and generalizations, statements of inferiority, expressions of contempt, dismissal or disgust, segregation or exclusion "in the form of calls for action, statements of intent, aspirational or conditional statements." As to the speech that requires additional analysis, it is the "[c]ontent attacking concepts, institutions, ideas, practices, or beliefs associated with protected characteristics, which are likely to contribute to imminent physical harm, intimidation or discrimination against the people associated with that protected characteristic." With respect to "Violence and Incitement", automatically banned content is the one, which could lead to death, admits past violence, promotes means and opportunities to kill or injure people and admits statements or intents to violence. Moreover, additional analysis is needed for the

[194] 'Hate Speech' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/> accessed 31 August 2022.

[195] 'Violence And Incitement' (*Meta*) <https://transparency.fb.com/policies/community-standards/violence-incitement/#policy-details> accessed 31 August 2022.

aforementioned statements in relation to enforcement agents and people accused of a crime, and where no specific means of doing harm are mentioned, but violence can still be a possible outcome.

Moving to the reports, the regular moderation information is published only concerning the "Hate Speech" policy, with a highly detailed breakdown for Facebook and Instagram separately. The categories of the report include prevalence, content actioned, proactive rate, appealed content, and restored content. The dynamics indicate that for the last several years the activity of Meta's moderation activities is rising, while users become less active in such activities as reporting and generally witness less hateful content. As to "Violence and Incitement", the statistics for it are not mentioned in a separate report and cannot be found on the policy page.

III.A.1.b. Actions during the Russian-Ukrainian war

In the context of the Russian-Ukrainian war, the company took some direct steps to vitiate the negative effects of incitement to violence against Ukrainians on its platform. These actions can be categorized into three groups: 1) focused on limiting the powers of actions sponsored by the aggressor state, 2) concentrated on stopping users from sharing risky content, and 3) protecting the safety of Ukrainian users. As for the first group, right after the beginning of the war, on February 25, the company announced the prohibition of Russian state governmental media from advertising and monetization capabilities.[196] In March it took even further steps, including the lower prioritization of posts from Russian state-sponsored media and the posts that had links to them, or Russian state websites.[197] This decision was enforced both on Facebook and Instagram, including usual posts and stories. Moreover, the posts of Russian state media were marked as such to notify the users of their nature.[198] After the request from "a number of governments", Meta restricted access to such prominent Russian propaganda news platforms as Russia Today and Sputnik both in the EU and the UK.[199] Finally, it banned

[196] (*Twitter*) <https://twitter.com/ngleicher/status/1497417241947607043> accessed 31 August 2022.

[197] Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022.

[198] Ibid.

[199] (*Twitter*) <https://twitter.com/nickclegg/status/1498395147536527360> accessed 31 August 2022; 'Meta's Ongoing Efforts Regarding Russia's Invasion Of Ukraine' (*Meta*, 2022)

a network of Russian-sponsored accounts that were responsible for coordinated inauthentic behavior (created fake accounts that spread fake news, hate speech, and incitement to violence) targeting the Ukrainian population.[200] Moving to the second group, Meta decided to warn users "when they try to share some war-related images that our systems detect are over one year old so people have more information about outdated or misleading images that could be taken out of context."[201] In addition, they went on to ban pages, groups, and accounts that share false information on multiple occasions.[202] Concerning the last group, it enhanced the privacy settings and tools, introduced encrypted messaging, and adjusted the management of private data.[203]

III.A.1.c. Mechanism of user reports

Having highlighted the company's policy nuances and proactive actions, we can analyze the operational aspects that empower users to act against infringing content. Meta divides the user experience in this respect into four stages: 1) reporting, 2) post-report communication, 3) takedown experience, and 4) warning screens. On the stage of reporting the user is able to pass on information about the post, which violates Meta's policies, choosing a particular policy concern.[204] Afterward, in the post-report communication stage, the user receives a notification as to the review team's decision and a link to the Support Inbox with further details.[205] It is also possible to appeal the result in case of dissatisfaction. The takedown experience stage concerns the perspective of the user, whose post was reported. In this regard, Meta communicates the reporting occasion with an explanation of the relevant policy and gives the user two options: to

<https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/> accessed 31 August 2022.

[200] 'Updates on Our Security Work in Ukraine' (*Meta*, 2022) <https://about.fb.com/news/2022/02/security-updates-ukraine/> accessed 31 August 2022.

[201] 'Meta's Ongoing Efforts Regarding Russia's Invasion Of Ukraine' (*Meta*, 2022) <https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/> accessed 31 August 2022.

[202] Ibid.

[203] Ibid.; Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022.

[204] 'How to Report Things' (*Facebook*) <https://www.facebook.com/help/reportlinks/> accessed 31 August 2022.

[205] 'Support Inbox' (*Facebook*) <https://www.facebook.com/support> accessed 31 August 2022.

agree or disagree with a decision. The warning screen stage is used for ordinary users when the content was decided to be of explicit nature or with subjective context and it is up to the user's consideration whether to push the button and familiarize with the content. Another interesting feature is the ability for users from Germany and Austria to submit reports on the ground of NetzDG and CPA, respectively.[206] The reports lodged both via Facebook and Instagram will be firstly reviewed according to Meta's Community Guidelines and deleted internationally, if it violates them. If the content will be in line with Meta's standards, the review team will then analyze them according to German or Austrian law and ban them in these countries, if the content is found to be illegal. Overall, Meta provides quite good instruments to engage users in the reporting process, provides an appeal function in case of disagreement, and even mentions the availability of a review by their multilingual team to ensure the objective analysis of content.[207]

III.A.1.d. Availability and reaction to inciting content

Despite the sophisticated descriptions of the review process, the platform did not delete or geoblock the pages, which are used for Russian propaganda and promote the war in Ukraine.[208] Furthermore, I managed to find content with the signs of incitement to violence against Ukrainians. For instance, both on Facebook and Instagram, it is possible to find content with uses of racial slurs toward Ukrainians in a hateful manner (such as "hohol", "banderovets", "zhidobanders").[209] Moreover, some posts compare Ukrainian people and soldiers to "fascists", "Nazis", "pigs" and "pedophiles."[210] Some content contains direct calls to kill Ukrainians or

[206] 'Austria Communication Platform Act' (*Facebook*) <https://www.facebook.com/help/3846278558774584/?helpref=hc_fnav> accessed 31 August 2022; 'What's The Difference Between CPA And The Instagram Community Guidelines?' (*Instagram*) <https://help.instagram.com/428536715033518/?helpref=related_articles> accessed 31 August 2022; 'What's The Difference Between NetzDG And The Instagram Community Guidelines?' (*Instagram*) <https://help.instagram.com/1787585044668150/?helpref=related_articles> accessed 31 August 2022.

[207] 'Report Something' (*Facebook*) <https://www.facebook.com/help/263149623790594?ref=tc> accessed 31 August 2022.

[208] (*Facebook*) <https://www.facebook.com/RusovDvizhenie> accessed 31 August 2022; (*Facebook*) <https://www.facebook.com/RussianEmbassy> accessed 31 August 2022; (*Facebook*) <https://www.facebook.com/groups/mid.dnr> accessed 31 August 2022.

[209] Examples of such posts are available on the Google Drive folder in case the posts get deleted. In the "Details" section such information can be found: relevant translations, explanations of context and racial slurs, and the links to the posts: 'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022, Facebook, Instagram.

[210] Ibid.

approve of their murder, using such phrases as "burn in hell", "death to Ukraine", mentions of extermination, physical mutilation, and injuries.[211] Another aspect is the denial of the territorial integrity of Ukraine, the use of chauvinistic Russian names for Ukrainian territories, such as "Malorossia" ("Little Russia)" or "Novorossia" ("New Russia"), and disrespect to Ukrainian national symbols.[212] Some posts try to argue that Ukrainian territories are Russian and will be soon captured by the Russian Army.[213] The aforementioned posts were reported to the review team for violating Meta's "Violence and Incitement" policy, as they contain infringing elements pursuant to it. The reports were made both under the CPA form (as the content was available to me in Austria during the writing process), and through regular reporting function. As a result, out of 11 posts only 5 of them were deleted (2 on Facebook and 3 on Instagram). The rest of the posts that contain infringing elements still can be reached on the platforms.[214]

III.A.2. Twitter

III.A.2.a. Scope of policies and the absence of reports

Moving to the next Big Tech company, which plays a prominent role in the ongoing Russian-Ukrainian war, Twitter also has exhaustive definitions of freedom of expression and incitement against protected groups and provides a decent reporting mechanism.

With regards to the differentiation of infringing tweets that fall out of the scope of freedom of expression, the company automatically bans the content with an intent to 1) "incite fear or spread fearful stereotypes about a protected category, including asserting that members of a protected category are more likely to take part in dangerous or illegal activities", 2) "incite others to harass members of a protected category on or off-platform", and 3) "incite others to discriminate in the form of denial of support to the economic enterprise of an individual or

[211] Ibid.

[212] Ibid.

[213] Ibid.

[214] (*Facebook*) <https://www.facebook.com/photo.php?fbid=117088227751551&set=pb.100083511233390.-2207520000..&type=3> accessed 31 August 2022; (*Facebook*) <https://www.facebook.com/RusovDvizhenie/photos/145264214782351> accessed 31 August 2022; (*Facebook*) <https://www.facebook.com/RusovDvizhenie/posts/pfbid02MdwgVgHhBjMyCBmnV3ujx17AWBB1v8asGWmvb9FZTaHRz6ErXvGSmGF49k8SKPc5l> accessed 31 August 2022; (*Instagram*) <https://www.instagram.com/p/ChfQseBoEGE/> accessed 31 August 2022; (*Instagram*) <https://www.instagram.com/p/CeW3nk3jsJP/> accessed 31 August 2022; (*Instagram*) <https://www.instagram.com/p/Cek-gMUsBMJ/> accessed 31 August 2022.

group because of their perceived membership in a protected category."[215] The company underlines that such content is prohibited under such conditions as "Wishing, hoping, or calling for serious harm on a person or groups of people" and should be reported as such.[216] Twitter also draws a line between truly infringing content and one with abusive or violent intent, such as the tweets with hyperbolized speech and exchanges between friends.[217] However, the company does not publish a report concerning the enforcement of this policy and the moderation activities pursuant to it.

III.A.2.b. Twitter's stance during the war

Concerning the proactive actions to protect Ukrainians, Twitter took actions that can be divided into two groups: 1) restriction of access for Russian and Belorussian state-sponsored media or the supporters of current Russian politics and 2) protection of Ukrainian users. As to the first group, it decided not to recommend or promote Russian and Belorussian state-led media and will label them as such.[218] The posts that contain links to them will be destined to the same fate.[219] Moreover, the company banned the aforementioned propaganda sources Russia Today and Sputnik in the EU,[220] and deleted more than 100 accounts that tweeted "#IStandWithPutin" for "coordinated inauthentic behavior."[221] With regards to the second group, Twitter enhanced the privacy policy availability for the region of Eastern Europe and

[215] 'Hateful Conduct Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy> accessed 31 August 2022.

[216] Ibid.

[217] 'Violent Threats Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/violent-threats-glorification> accessed 31 August 2022.

[218] Samuel Benson, 'Twitter To Label All State-Affiliated Russia Media' <https://www.politico.com/news/2022/02/28/twitter-label-state-affiliated-russia-media-00012351> accessed 31 August 2022; 'About Government And State-Affiliated Media Account Labels On Twitter' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/state-affiliated> accessed 31 August 2022; Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022.

[219] Ibid.

[220] Elizabeth Culliford, 'Twitter To Comply With EU Sanctions On Russian State Media' *Reuters* (2022) <https://www.reuters.com/technology/twitter-comply-with-eu-sanctions-russian-state-media-2022-03-02/> accessed 31 August 2022.

[221] Ben Collins, and Natasha Korecki, 'Twitter Bans Over 100 Accounts That Pushed #IStandWithPutin' *CBS News* (2022) <https://www.nbcnews.com/tech/internet/twitter-bans-100-accounts-pushed-istandwithputin-rcna18655> accessed 31 August 2022.

started to monitor "vulnerable high profile accounts" of journalists, state officials and activists against targeting or manipulation.[222]

III.A.2.c. Reporting instruments for users

The users of this social network can also avail themselves of the opportunity to report the content. The most basic option would be to report the tweet directly, choosing the reason "It's abusive or harmful" and selecting "Threatening violence or physical harm."[223] Afterward, the users can choose their relation to the protected group and select up to five tweets for reporting. In addition, a specific abusive behavior reporting form is available, where the users can select the infringing account, up to five tweets that violate the policy, and give additional details as to the contents' nature and context.[224] The reports can also be followed by a notification from Twitter's review team if the user chooses this option.

III.A.2.d. The presence of hateful content and reaction to it

Like with Facebook, the highly developed method of policing the content and enabling users to report it does not mean that the platform is flawless in its counteraction to Russian state-led propaganda. Twitter failed to get rid of pages that regularly spread fake news and mislead users using common methods of propaganda.[225] This social network is also not free from incitement to violence against protected groups. For example, some tweets have racial slurs towards the Ukrainians such as "hohols" and "Ukronazis."[226] The others incite to death

---

[222] (*Twitter*) <https://twitter.com/TwitterSafety/status/1497353968565075968?s=20&t=VwP5VUPj2QuDCgZhN9az2w> accessed 31 August 2022.

[223] 'Violent Threats Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/violent-threats-glorification> accessed 31 August 2022.

[224] 'Staying Safe On Twitter And Sensitive Content' (*Twitter*) <https://help.twitter.com/en/forms/safety-and-sensitive-content/abuse/legal-rep> accessed 31 August 2022.

[225] James Clayton, 'How Kremlin Accounts Manipulate Twitter' *BBC* (2022) <https://www.bbc.com/news/technology-60790821> accessed 31 August 2022; Timothy Graham and Jay Daniel Thompson, 'Russian Government Accounts Are Using A Twitter Loophole To Spread Disinformation' *The Conversation* (2022) <https://theconversation.com/russian-government-accounts-are-using-a-twitter-loophole-to-spread-disinformation-178001> accessed 31 August 2022; (*Twitter*) <https://twitter.com/RussianEmbassy> accessed 31 August 2022; (*Twitter*) <https://twitter.com/M_Simonyan> accessed 31 August 2022.

[226] Examples of such posts are available on the Google Drive folder in case the posts get deleted. In the "Details" section such information can be found: relevant translations, explanations of context and racial slurs, and the links to the posts: available at: 'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022, Twitter.

and violent actions with such phrases as “death to hohols” and “all fucking Ukronazis come on order to exterminate.” Finally, some posts diminish the territorial integrity of Ukraine and justify the past genocide of the Ukrainian people. All of these posts fall into the category of harmful content pursuant to Twitter’s policy. I reported the aforementioned tweets and out of five posts, three were deleted. The two posts that contain racial slurs and were not banned still can be found on the platform.[227]

III.A.3. YouTube

III.A.3.a. Policy implementation and regular reporting

YouTube is another example of a Big Tech content platform that provides extensive explanations of its understanding of human rights protection and violations in its Community Guidelines. Under the company’s standards, incitement to violence is included in the “Hate speech” policy by prohibiting the promotion of violence or hatred against protected groups with such attributes as ethnicity, nationality, and race.[228] If the content contains encouragement to violence or incites hatred against individuals or protected groups based on the above-mentioned attributes, it gets automatically banned. However, YouTube allows the use of themes relating to these violations, if it is mentioned for educational, documentary, scientific, or artistic purposes and does not promote hate speech.[229]

As to reporting activities, YouTube regularly provides sufficient information on its actions, by issuing schemes, diagrams, and descriptions that contain elucidating breakdowns according to the policy that has been triggered during the review.[230] YouTube’s statistics include such categories as removed videos, comments, and channels. The information is available in numeric form (total number and percentage), removal reason, views,

[227] (*Twitter*) <https://twitter.com/VladPunished/status/1561123899403894784> accessed 31 August 2022; (*Twitter*) <https://twitter.com/MirelaBjelic/status/486245763669053440> accessed 31 August 2022.

[228] 'Hate Speech Policy' (YouTube Help) <https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436> accessed 31 August 2022.

[229] Ibid.

[230] 'YouTube Community Guidelines Enforcement' (*Google Transparency Report*) <https://transparencyreport.google.com/youtube-policy/removals> accessed 31 August 2022.

country/region, and source of the first detection. Moreover, in the statistics dedicated to hate speech, YouTube enlists several examples of its proceeding with respect to this policy.[231]

III.A.3.b. Proactive actions in times of war

Throughout the course of the Russian-Ukrainian war, YouTube has been quite proactive in its actions with such purposes as 1) banning and limiting Russian content targeting Ukrainians and 2) taking precautions for the sake of the Ukrainian people. Concerning the first purpose, it started to label and block the channels, associated with the Russian state both in Europe (including Ukraine) and globally.[232] Moreover, the company stopped recommending content of Russian-sponsored content and suspended monetization for Russian content generated by the views of Russian users.[233] On March 11 YouTube also stated that it removed more than 1,000 channels and 15,000 videos that violated its policies, especially in the context of targeting Ukrainians and justifying the actions of the Russian army.[234] With respect to the second aim, YouTube enhanced security measures and advanced protection for Ukrainian users in addition to other numerous actions done by Google on its other platforms, including Google Maps and Gmail.[235]

---

231 'Featured Policies – Hate Speech' (*Google Transparency Report*) <https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=en> accessed 31 August 2022.

232 (*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335085122666500> accessed 31 August 2022; (*Twitter*) <https://twitter.com/YouTubeInsider/status/1498772480034365440> accessed 31 August 2022; Paresh Dave, 'Google Blocks RT, Other Russian Channels From Earning Ad Dollars' *Reuters* (2022) <https://www.reuters.com/technology/youtube-blocks-rt-other-russian-channels-generating-revenue-2022-02-26/> accessed 31 August 2022.

233 (*Twitter*) <https://twitter.com/YouTubeInsider/status/1498772481309437952?s=20&t=SQJVgNzHU6C1Hn6PGUI8rg> accessed 31 August 2022; Matt Novak, 'YouTube Stops Monetization For Video Creators In Russia But There's One Exception' <https://gizmodo.com/youtube-stops-monetization-for-video-creators-in-russia-1848633335> accessed 31 August 2022.

234 (*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335030168899595> accessed 31 August 2022; (*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335119914381314?s=20&t=j_NJ6tVDMG4MQOPjRSxelA> accessed 31 August 2022.

235 Kent Walker, 'Helping Ukraine' <https://blog.google/inside-google/company-announcements/helping-ukraine/?utm_source=tw&utm_medium=social&utm_campaign=og&utm_content=&utm_term> accessed 31 August 2022; Marc Cieslak and Tom Gerken, 'Ukraine Crisis: Google Maps Live Traffic Data Turned Off In Country' *BBC* (2022) <https://www.bbc.com/news/technology-60561089> accessed 31 August 2022.

III.A.3.c. Compliance of incitements' policing activities and user reports

Moving to the reporting remedies, available to YouTube users, they can report virtually any content, present on the platform, including channels, videos, comments, thumbnails, playlists, links, ads, and live chat messages.[236] To do that the user simply has to click the report button and select a policy reason for the claim. The largest space for imagination is given to the user in video reports, where the argumentation can be also inserted. Moreover, if a user thinks that a certain video is illegal, a legal report can be lodged with the specific selection of the user's country of origin, link and description of the violated law, and the description of the violation.[237] The legal reporting option allows users to submit claims on the basis of any national or international law, including CPA and NetzDG. For the latter laws, YouTube even provides specific pages, where they are explained in detail and YouTube's modus operandi in relation to these laws is described.[238] Also, the NetzDG page has a report on taken actions, examples of cases, and a separate reporting form with the selection of NetzDG violations and the space to mention the time code of the video, where the infringement occurs.[239]

Comparing YouTube to other content platforms, it would probably have the most detailed and sophisticated policy definitions and reporting instruments. However, the platform did not ban or flag some channels, which regularly publish Russian propaganda and undermine the validity of proven historical facts.[240] As to the hateful content, there are almost no videos

---

[236] 'Report Inappropriate Videos, Channels, And Other Content On YouTube' (*YouTube Help*) <https://support.google.com/youtube/answer/2802027?hl=en&co=GENIE.Platform%3DDesktop&oco=0#zippy=%2Cprivacy-reporting%2Clegal-reporting%2Creport-a-video%2Creport-a-channel%2Creport-a-playlist%2Creport-a-thumbnail%2Creport-a-comment%2Creport-a-link%2Creport-a-live-chat-message%2Creport-an-ad (Last accessed: 25 August 2022)> accessed 31 August 2022.

[237] 'Other Legal Complaint' (*YouTube Help*) <https://support.google.com/youtube/contact/other_legal> accessed 31 August 2022.

[238] 'Removals Under The Network Enforcement Law' (*YouTube Transparency Report*) <https://transparencyreport.google.com/netzdg/youtube?hl=us> accessed 31 August 2022; 'Text-Based Legal Complaints Under Kopl-G' (*YouTube Help*) <https://support.google.com/youtube/answer/11052657?hl=en&ref_topic=6154211> accessed 31 August 2022.

[239] 'Removals Under The Network Enforcement Law' (*YouTube Transparency Report*) <https://transparencyreport.google.com/netzdg/youtube?hl=us> accessed 31 August 2022; 'Report content under the Network Enforcement Law' (YouTube Help) <https://support.google.com/youtube/contact/netzdg> accessed 31 August 2022.

[240] Ivan Makridin, 'Banned' Russian Propaganda Still Easy To Find On YouTube' <https://www.codastory.com/newsletters/russian-propaganda-youtube/> accessed 31 August 2022; Esteban Ponce De León, 'Russian Propaganda Is Thriving In Spanish' <https://slate.com/technology/2022/06/kremlin-propaganda-spanish-latam.html> accessed 31 August 2022; 'Anatolii Sharii's YouTube Channel' (*YouTube*) <https://www.youtube.com/user/SuperSharij> accessed 31 August 2022.

on the platform that contain incitement to genocide or hate speech against Ukrainians. However, some mentions of violence toward Ukrainian people can be found in the comment section, where some users mention their desire to exterminate Ukrainian nationalists.[241] The report on this information has been lodged to YouTube's review team, but no action was taken by the platform.

III.A.4. TikTok

III.A.4.a. Policy scope and actions during the war

In the case of TikTok, freedom of expression does not find its place in the social network's policies, contrary to incitement to violence, which is best mirrored in such policies as "Hateful behavior" and "Violent extremism." TikTok states that it bans content that encourages people to commit violence, including statements, imagery, and calls that "intend to inflict physical injuries on an individual or a group" and "encourage others to commit or that advocate for violence."[242] The platform does not provide any exclusions or limitations from this policy.

With respect to reporting activities, TikTok does not issue general information concerning content moderation, but provided specific information in the context of the Russian-Ukrainian war, mentioning the removal of 41,191 videos with harmful information, deletion of 321,784 unauthentic Russian and 46,298 Ukrainian accounts, which had 343,961 videos fake news videos, labeled 49 Russian state-sponsored accounts and deleted 6 networks and 204 accounts that acted in conformity to influence the users and mislead them.[243]

In addition to the aforementioned activities, the company took a few additional steps to protect Ukrainian users and block the negative actions of Russian users and state-backed

[241] Examples of such posts are available on the Google Drive folder in case the posts get deleted. In the "Details" section such information can be found: relevant translations, explanations of context and racial slurs, and the links to the posts: 'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022, YouTube.

[242] 'Violent Extremism' (*TikTok*) <https://www.tiktok.com/community-guidelines#39> accessed 31 August 2022.

[243] 'Bringing more context to content on TikTok' (*TikTok*) <https://newsroom.tiktok.com/en-us/bringing-more-context-to-content-on-tiktok?utm_source=COMMSTWITTER&utm_medium=social&utm_campaign=030622> accessed 31 August 2022.

profiles. For instance, they managed to geoblock Russian state media, including Russia Today and Sputnik,[244] and added tips for enhancing users' literacy concerning their online behavior.[245]

III.A.4.b. Availability of reports to users and disregard to them

Concerning the reporting functionality, users are able to report videos, comments, direct messages, users, sounds, and hashtags.[246] To fulfill this, it is enough to select a report option on the screen and follow onscreen instructions, providing the description of the issue. Alternatively, it is possible to submit a complaint via the online form, where it is possible to choose an issue, explain the rationale for reporting and attach files for the evidentiary basis.[247]

Despite the described steps and policy considerations, Russian propaganda still has its place on the platform via the presence of pages that spread fake news and calls in favor of the war.[248] Besides, it is still possible to find hateful content that targets the Ukrainian population. For instance, some videos promote the killings of Ukrainians and use racial slurs, which end up in phrases like "death to banderovtsy."[249] Moreover, the content contains the discussions of the occupation of Ukraine by Russian state officials, the claim that the Ukrainian Kherson oblast belongs to Russia, and the use of chauvinistic terms for Ukrainian territories, such as

[244] (*Twitter*) <https://twitter.com/BobbyAllyn/status/1498421323135012865> accessed 31 August 2022.

[245] 'Bringing more context to content on TikTok' (*TikTok*) <https://newsroom.tiktok.com/en-us/bringing-more-context-to-content-on-tiktok?utm_source=COMMSTWITTER&utm_medium=social&utm_campaign=030622> accessed 31 August 2022; Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022.

[246] 'Report a problem' (*TikTok*) <https://support.tiktok.com/en/safety-hc/report-a-problem> accessed 31 August 2022.

[247] 'Share your feedback' (*TikTok*) <https://www.tiktok.com/legal/report/feedback> accessed 31 August 2022.

[248] David Gilbert, 'Russian Tiktok Influencers Are Being Paid To Spread Kremlin Propaganda' Vice (2022) <https://www.vice.com/en/article/epxken/russian-tiktok-influencers-paid-propaganda> accessed 31 August 2022; David Shepardson and Echo Wang, 'U.S. Senators Press Tiktok On Whether It Allows Russian "Pro-War Propaganda"' *Reuters* (2022) <https://www.reuters.com/business/media-telecom/us-senators-press-tiktok-whether-it-allows-russian-pro-war-propaganda-2022-06-17/> accessed 31 August 2022; Euronews, 'Pro-Kremlin Propaganda Still Rife On Tiktok Despite Ban' (2022) <https://www.euronews.com/my-europe/2022/04/13/pro-kremlin-propaganda-still-rife-on-tiktok-despite-ban-new-report-shows> accessed 31 August 2022; (*TikTok*) <https://www.tiktok.com/@russian_military_guy> accessed 31 August 2022; (*TikTok*) <https://www.tiktok.com/@russian_mma> accessed 31 August 2022.

[249] Examples of such posts are available on the Google Drive folder in case the posts get deleted. In the "Details" section such information can be found: relevant translations, explanations of context and racial slurs, and the links to the posts: 'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022, TikTok.

"Malorossia."[250] In addition, some videos contain the diminishing use of Ukrainian national attributes, like the burning of Ukrainian passport.[251] All of these posts contain infringing elements under TikTok's policies and were reported to their moderation team. However, TikTok decided not to delete the posts and they are still available to the general public.

III.A.5. Telegram

III.A.5.a. Indifference to definitions in policies and limited proactive steps

Moving to the last social network that is abundantly used during the current war, Telegram has scarce mentions of freedom of expression and does not include any remarks concerning incitement to violence.[252] Furthermore, the company does not publish any reports as to the moderation activities, apart from the regular updates, provided on the channel "ISIS watch", which informs the users about the deletion of terrorist channels and bots.[253]

As to the proactive actions throughout the war's course, Telegram only agreed to geoblock Russia Today and Sputnik in the EU after the requests from European governments.[254]

III.A.5.b. Narrow functions of user reports and abundance of hateful content

Telegram users are not able to report specific content to the social network's team. However, there is an option to report sticker sets, channels, and bots either by pushing a "report" button, choosing the reason for the report (includes "Violence"), and picking the infringing posts or by sending an email to abuse@telegram.org with further details.[255]

Taking into consideration the volatile nature of Telegram's policy and the absence of any mentions of incitement to violence, the whole framework of content moderation seems weaker in the case of this social network. The amount of hateful content on Telegram is much greater compared to all the previous platforms altogether and includes not only hateful messages but also graphic imagery with racial slurs and direct calls to commit massive atrocities

[250] Ibid.

[251] Ibid.

[252] 'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022.

[253] 'ISIS Watch' (*Telegram*) <https://t.me/isiswatch> accessed 31 August 2022.

[254] Mark Scott, 'Telegram Bans Russian State Media After Pressure From Europe' <https://www.politico.eu/article/russia-rt-media-telegram-ukraine/> accessed 31 August 2022.

[255] 'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022.

against the Ukrainian population. Telegram failed to ban Russian state-supported channels or even indicate the highly subjective nature of their content.[256] Moreover, the platform did not ban channels that have racial slurs and incitement to violence both in their posts and names.[257] Despite the inactivity of Telegram in this respect, I sent the reports to the aforementioned channels with a list of posts with hateful content. Although the channels' names and their posts were of infringing nature, Telegram's moderation team did not react to the reports.

III.B. Lessons

The difference in the apprehension of incitement to violence became evident as we went through the nuances of moderation activities across six different platforms that are most actively used in the course of the Russian-Ukrainian war. The strongest response to the online hostilities was demonstrated by the social networks domiciled in the US (Meta's Facebook and Instagram, Twitter, and YouTube), contrary to TikTok and Telegram, which are based in China and the UAE, respectively. As to the geographical scope, the strongest response was seen in Europe and the EU, in particular, which can be explained by existing and forthcoming legislation, government-imposed sanctions, and official requests to social networks. Due to the various responses and uneven efficiency of the social networks' actions, it is necessary to establish the patterns that influenced the behavior of the companies. This information is highly important to give the states, social networks, public sector, and users useful recommendations that may help enhance the reactive framework toward digital incitement to violence in the case of ongoing Russian-Ukrainian war and the potential situations that may arise in the future.

III.B.1. States

Modern states are the most authoritative actors in the scheme of multistakeholder content governance due to the concentration of legislative, regulatory, and enforcement roles. The actions of state bodies have the power to create certain business climates, around which companies build their compliance mechanisms and choose specific modes of action. As we

[256] (*Telegram*) <https://t.me/margaritasimonyan> accessed 31 August 2022; (*Telegram*) <https://t.me/vysokygovorit> accessed 31 August 2022; (*Telegram*) <https://t.me/vladlentatarsky> accessed 31 August 2022.

[257] (*Telegram*) <https://t.me/hohlopidorislavarossii> accessed 31 August 2022; (*Telegram*) <https://t.me/trupvsy1> accessed 31 August 2022; (*Telegram*) <https://t.me/ukropi> accessed 31 August 2022; (*Telegram*) <https://t.me/otborsalo> accessed 31 August 2022; (*Telegram*) <https://t.me/Hohli_pidorasi> accessed 31 August 2022.

have seen in the passages concerning state recognition of freedom of expression and incitement to violence, European countries tend to adopt stringent norms that are of utmost importance to the legal sphere. With the adoption of authoritative laws and active judicial enforcement of them, it is possible to ensure that all the actors in the field will act with the necessary respect and caution concerning the protected rules.

Judging from the lessons of Germany and Austria, the first lesson for the states is the necessity of an existent legislative framework for the digital sphere with solid enforcement and monitoring systems. Such acts as NetzDG and CPA gained considerable authority and eventually became a part of compliance procedures of such Big Tech companies as Meta and Google, which proves the positive influence of state regulation on business operations. The EU has already acknowledged the importance of such steps, which led to the adoption of the DSA, which is soon to bring the novelties of German and Austrian legislation with some adjustments and modifications on the EU-wide level. Even despite the flaws of these legislative act, they are an important step to combat digital hateful content. Other political players should also expand the regular norms concerning human rights and incitement to the online sphere to ensure that users from their countries are protected from potential online abuses and hazards, and empower ordinary citizens to stand up to hostilities. It is especially relevant for the countries, which are already suffering from digital incitements to violence, but have not yet introduced any legislative solutions to such problems.

Another recommendation is the acknowledgment of international standards and their wide implementation by national institutions. One of the most prominent examples is the enforcement of UNGP norms concerning national action plans that urges businesses to be compliant with human rights standards. In addition to action plans, states must underline the importance of multistakeholder cooperation in their actions and legislative pieces. Judging from the EU's example, the introduction of national action plans and approval of multistakeholder governance in the text of DSA helped to ensure that social networks would respect the norms, adopted in the EU. Moreover, it is highly important to introduce legislation that enforces international interpretation of freedom of expression, incitement to violence, and their balancing, as the uniform response to the violation of these standards will guarantee that there are no safe havens for non-compliant actors.

In all the aforementioned activities it is also obligatory for states to take into consideration the opinions of users and representatives of the social sector to ensure that their decisions represent the social attitude and academic state of the art.

Finally, sometimes it is necessary for the states to directly address the companies on relevant topics and find ways for cooperation between the actors. Judging from the EU and Ukraine's examples, the inquiry to social networks can rapidly result in cooperation from social networks due to the mere authority of state representatives. However, such steps should also be done with due diligence to human rights and mark the safety and integrity of each and every citizen as a number one priority to vitiate the possibility of oppressive exploitation of state's policing functions.

III.B.2. Social networks

With respect to the intermediaries, they have the most effective operational position to act against digital incitement to violence, so their actions are extremely important for the sake of protected groups.

First of all, social networks should act according to the legal requirements of their users' countries of origin. Due to the complexity of the cross-border online sphere, they should not limit themselves to the legal spectrum of the state, where their headquarters are placed. Big Tech companies should mirror both international and national legal standards in their policies and treat their users, pursuant to them, even if some users' countries have not yet introduced profound legislation.

Apart from that, businesses should not treat legal norms as the ceiling of their operations, but as a starting point for their proactive actions. In this sense, it is important to adopt the standards of CSR and be the trendsetter in the field of content moderation and compliance.

III.B.3. Users and social sector

Moving to the groups, which are the most numerous but have the least degree of authority, it is necessary to remember that collective efforts can lead to extremely powerful results and finally draw the attention of international and regional bodies, states, and social networks to the relevant issues that require some action.

The representatives of the social sector should ensure that actual problems fall into the scope of their professional activities and that their opinions are heard by the actors in power.

Whether these are the protests of social activists or the academic works of reputable scholars, they should adhere to the needs of societies and draw the lines between right and wrong.

Finally, the users should remember that their role in the multistakeholder system of governance is not limited to the mere recipients of the legal norms and business policies. The activities of users, both online and offline can guide the actions of other players and become too hard to disregard. As the current war indicated, users' boycotts and calls to action played a role no less important than that of the social sector or states.

CONCLUSION

The power of the Internet unfolds with an unprecedented speed, unleashing its benefits and hazards in times of extreme events. The horrifying Russian-Ukrainian war keeps demonstrating how Web 2.0 platforms can contribute to the rapid spread of important information and simultaneously be exploited as a malicious instrument of human rights violations. Incitement to violence is one of those actions, which modern online users had a chance to familiarize themselves with. Taking into consideration the regular historic occurrence of this inchoate crime, I embarked on a research journey to analyze the existent legislative answers to digital incitement to violence, and how social networks implement them in the context of Russian-Ukrainian war.

The decision was made to use the theoretical background of internet governance, which sufficiently explains the interaction of online actors. To tailor, this interplay to the topic of content moderation, I chose the approach of multistakeholder content governance, which describes the give-and-takes between the relevant actors. With the aim of finding the most appropriate strategy for the deployment of cooperation between states, social networks, and, to the lesser degree, the social sector and users, I focused on the mechanism of intermediary liability, which places the burden of policing activities on social networks. This approach is chosen by the EU legislators and found its place in such legislative acts, as NetzDG, CPA, and DSA.

Having established the most lucrative approach that can be exploited by the actors of multistakeholder content governance, I enlisted the framework of rules that should be adhered to in the context of combating digital incitement to violence. Such framework includes business respect for human rights (UNGP), protection of freedom of expression (UDHR, ICCPR), and prohibition of incitement to violence (ICCPR, decisions of international criminal tribunals). The regional and national implementation of these norms proved to be highly uneven. Even the majority of European countries, which had the legislation, compliant with international norms, failed to share its effects to the online sphere. As a result, the only European legislative documents that explicitly address the issue of digital incitement to violence are the aforementioned NetzDG, CPA, and DSA. Despite the abundant critique, these legislative acts serve as a powerful means of counteraction to hateful content by establishing the policing scope of actions for the social networks.

Assessing the actions of the most prominent social networks, which are used during the course of the Russian-Ukrainian war, I analyzed the implementation of the aforementioned rules and the compliance of moderation activities of such platforms as Facebook, Instagram, Twitter, YouTube, TikTok, and Telegram. As to policy considerations, Facebook, Instagram, Twitter, and YouTube proved to be the most compliant, since their policy had relevant mentions of international documents and explicit definitions of their commitments to human rights and freedom of expression. Moreover, they had specific policies regarding the moderation of content with incitement to violence and user reporting mechanisms. TikTok had considerations only as to hateful content, but also managed to introduce reporting functions. Telegram had the weakest policy framework with a brief mention of "free expression" as one of its values and a reporting mechanism that allowed users to pass on information about infringing channels. Reports concerning moderation activities were published by Facebook, Instagram, and YouTube, with the latter being the most informative. As a result of the search for infringing content on the platforms, YouTube turned out to be the most compliant with minor incitements in the comment sections. The other platforms contained posts with racial slurs, diminishing mentions of Ukrainian sovereignty and national symbols, and ordinary incitement to violence. Telegram had the most tenuous moderation mechanisms, in addition to numerous infringing posts, racial slurs and incitements that were present even in the names of the channels. Eventually, I lodged the reports for several posts on Facebook, Instagram, Twitter, TikTok, and Telegram that contained the infringing elements pursuant to the platforms' policies and legal standards. The first three platforms deleted half of the reported content, and the latter two did not react at all. In the end, the infringing content was left available on all of the platforms, including the comment section on YouTube.

With the aim of helping content governance stakeholders to combat incitement and other illegal content both during the course of the Russian-Ukrainian war and in future possible conflicts, I presented a number of recommendations for each actor. Namely, states are responsible for online-applicable legislation and its enforcement, with such examples as Germany (NetzDG), Austria (CPA), and the EU (DSA) being the most prominent followers of such advice. Their actions were among the reasons the moderation activities of social networks were the most diligent in Europe and mostly adhered to governmental requests. Social networks have to adopt the principles of CSR and act according to national and international legal standards, conducting risk assessment activities to exterminate the threats to their consumers.

The social sector has to find the spheres, which lack public highlight and actively scrutinize them, ensuring that their voices are heard by the people in power. Finally, users have to acknowledge the potential of their unified activities and act not only as recipients of rules and policies but also as advocates of necessary changes. Naturally, all of these recommendations are not an ultimate answer to digital illegal content and should be complemented by the elaboration of multidisciplinary research. However, they represent the bare minimum, which is required to continue the counteraction to online atrocities.

I strongly believe that the findings of this thesis can contribute to the fight against an ongoing outbreak of severe injustices that serve as a threat to the peaceful development of young democracies under the principle of rule of law. Despite the hardships of regulation in the plethora of online communications between millions of actors, it is possible to adopt the instruments that would secure the targeted groups and make social networks a safe place that would correspond to the initial aims of the Web's developers – "to serve humanity"[258] without the misuse "by those who want to exploit, divide and undermine."[259] With this in mind, I strongly call for further research in the spheres of Internet governance, content moderation, and efficient prevention of online crimes, with the hope that the digital sphere will have no place for the implementation of dystopian ideas.

[258] Doris Obermair, 'Tim Berners-Lee: The Goal Of The Worldwide Web Is To Serve Humanity' <https://www.instituteofnext.com/tim-berners-lee-the-goal-of-the-worldwide-web-is-to-serve-humanity/> accessed 31 August 2022.

[259] DW, 'Web Inventor Tim Berners-Lee Unveils Plan To Save The Internet' <https://www.dw.com/en/web-inventor-tim-berners-lee-unveils-plan-to-save-the-internet/a-51395985> accessed 31 August 2022.

GLOSSARY

Multistakeholder content governance	The constant elaboration and implementation of Internet-related principles, rules, mechanisms, and procedures in the field of content sharing and moderation by states, social networks, social sector (activists and academia), and users with interdependence between each other.
Internet governance	The development and application by governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures and programmes that shape the evolution and use of the internet.[260]
Web 2.0	The current state of the internet, which has more user-generated content and usability for end-users compared to its earlier incarnation, Web 1.0. In general, Web 2.0 refers to the 21st-century Internet applications that have transformed the digital era in the aftermath of the dotcom bubble.[261]

[260] Working Group on Internet Governance, 'Report Of The Working Group On Internet Governance' (2005) <http://www.wgig.org/docs/WGIGREPORT.pdf> accessed 31 August 2022, p. 4.

[261] William Kenton, 'Web 2.0' (*Investopedia*) <https://www.investopedia.com/terms/w/web-20.asp> accessed 31 August 2022.

BIBLIOGRAPHY

Legal Sources

United Nations Documents

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR)

OHCHR 'Guiding Principles on Business and Human Rights' (2011) UN Doc HR/PUB/11/04

UN Committee on the Elimination of Racial Discrimination, 'General Recommendation No. 32 on The Meaning and Scope of Special Measures in the International Convention on the Elimination of Racial Discrimination' (24 September 2009) UN Doc CERD/C/GC/32

UNGA 'Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred' (11 January 2013) UN Doc A/HRC/22/17/Add.4

UNGA Res 60/1 (24 October 2005) UN Doc A/RES/60/1

UNGA Rome Statute of the International Criminal Court (adopted 17 July 1998, last amended 2010) UN Treaty Series vol. 2187 no. 38544 (Rome Statute)

Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR)

UNSC Statute of the International Criminal Tribunal for Rwanda (adopted 8 November 1994 UNSC Res S/RES/955, last amended 2006) (ICTR)

UNSC Statute of the International Criminal Tribunal for the Former Yugoslavia (adopted 25 May 1993 UNSC Res 827/1993, last amended 17 May 2002) (ICTY)

Regional Bodies' Documents

African Union, Convention on Cyber Security and Personal Data Protection (Malabo, 27 June 2014)

Association of Southeast Asian Nations, ASEAN Human Rights Declaration (19 November 2012)

Council of Europe, Additional Protocol to the Convention on Cybercrime (Strasbourg, 28 January 2003) European Treaty Series - No. 189

Council of Europe, Convention on Cybercrime (Budapest, 23 November 2001) European Treaty Series - No. 185

Council of Europe, European Convention on Human Rights (Rome, 4 November 1950) (ECHR)

European Union, Charter of Fundamental Rights of the European Union (26 October 2012) 2012/C 326/02

Organization of African Unity, African (Banjul) Charter on Human and Peoples' Rights (27 June 1981) OAU Doc CAB/LEG/67/3

Organization of American States, American Convention on Human Rights "Pact Of San Jose, Costa Rica" (B-32)

Constitutions

Constitution of the Russian Federation

Constitution of Ukraine

U.S. Constitution, Amendment 1

Legislative Acts

18 U.S. Code § 1091 – Genocide

18 U.S. Code § 373 - Solicitation to commit a crime of violence

Brottsbalk (1962:700)

Bundesgesetz über Maßnahmen zum Schutz der Nutzer auf Kommunikationsplattformen (Kommunikationsplattformen-Gesetz – KoPl-G) 2020

Criminal Justice Act 2003

Explanatory Memorandum to DSA

Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG) 2017

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal

Loi 90-615 du 13 juillet 1990

Proposal for a Regulation on a Single Market For Digital Services (Digital Services Act) 2020

Proposal for a Regulation on Digital Markets Act (Digital Markets Act) 2020

Public Order Act 1986

Racial and Religious Hatred Act 2006 (England and Wales)

State Duma, Criminal Code of the Russian Federation

State Duma, Federal Law No. 31-FZ of 4 March 2022

Strafgesetzbuch 1975

Verkhovna Rada of Ukraine, Criminal Code of Ukraine

Cases

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

Magyar Tartalomszolgáltatók Egyesülete and Indexhu Zrt v Hungary App no 22947/13 (ECtHR, 2 February 2016)

Nahimana et al. (Media case) (Appeal Judgement) ICTR-99-52-A (28 November 2007)

Ngirabatware Augustin (Appeal Judgement) MICT-12-29-A (18 December 2014)

Prosecutor v. Ruggiu (Judgement and Sentence) ICTR-97-32-I (1 June 2000)

The Sunday Times v. United Kingdom App no 6538/74 (ECtHR 26 April 1979)

Zana v Turkey App no 18954/91 (ECtHR 25 November 1997)

Government Publications

European Commission, 'Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: A Renewed EU Strategy 2011–14 For Corporate Social Responsibility' (European Commission 2011)

European Commission, 'Green Paper Promoting A European Framework For Corporate Social Responsibility' (2001)

European Parliamentary Research Service, 'Potentially Negative Effects Of Internet Use' (European Parliament 2022)

Federal Bureau of Investigation, 'Internet Crime Report 2021' (Federal Bureau of Investigation 2021)

President of Ukraine, 'Decree №119/2021 “On The National Strategy For Human Rights”' (President of Ukraine 2021)

Reynders D, 'Countering Illegal Hate Speech Online: 6Th Evaluation Of The Code Of Conduct' (European Commission 2021)

'The EU Code Of Conduct On Countering Illegal Hate Speech Online' (*European Commission*) <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct> accessed 31 August 2022

Reports

Article 19, 'Responding To ‘Hate Speech’: Comparative Overview Of Six EU Countries' (Article 19 2018) <https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf> accessed 31 August 2022

Bukovska B, Callamard A, and Parrmer S, 'Towards An Interpretation Of Article 20 Of The ICCPR: Thresholds For The Prohibition Of Incitement To Hatred' (Article 19 2010)

<https://www2.ohchr.org/english/issues/opinion/articles1920_iccpr/docs/CRP7Callamard.pdf> accessed 31 August 2022

Buri I, and van Hoboken J, 'The Digital Services Act (DSA) Proposal: A Critical Overview' (Digital Services Act (DSA) Observatory, Institute for Information Law (IViR), University of Amsterdam 2021) <https://dsa-observatory.eu/wp-content/uploads/2021/11/Buri-Van-Hoboken-DSA-discussion-paper-Version-28_10_21.pdf> accessed 31 August 2022

Dibbets A, Sano H, and Zwamborn M, 'Indicators In The Field Of Democracy And Human Rights: Mapping Of Existing Approaches And Proposals In View Of Sida's Policy' (The Danish Institute for Human Rights 2010)

Global Network Initiative, 'Content Regulation And Human Rights' (Global Network Initiative 2020) <https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf> accessed 31 August 2022

Human Rights Watch, 'Russia, Ukraine, And Social Media And Messaging Apps: Questions And Answers On Platform Accountability And Human Rights Responsibilities' (Human Rights Watch 2022) <https://www.hrw.org/news/2022/03/16/russia-ukraine-and-social-media-and-messaging-apps> accessed 31 August 2022

International IDEA, 'Press Freedom And The Global State Of Democracy Indices' (The Global State of Democracy in Focus 2019)

Jahangir A, and Diène D, 'Report Of The Special Rapporteur On Freedom Of Religion Or Belief, Asma Jahangir, And The Special Rapporteur On Contemporary Forms Of Racism, Racial Discrimination, Xenophobia And Related Intolerance, Doudou Diène, Further To Human Rights Council Decision 1/107 On Incitement To Racial And Religious Hatred And The Promotion Of Tolerance' (UN Human Rights Council 2022) <https://digitallibrary.un.org/record/583355> accessed 31 August 2022

UNESCO, 'Keystones To Foster Inclusive Knowledge Societies: Access To Information And Knowledge, Freedom Of Expression, Privacy And Ethics On A Global Internet' (UNESCO 2015) <https://unesdoc.unesco.org/ark:/48223/pf0000232563> accessed 31 August 2022

Working Group on Internet Governance, 'Report Of The Working Group On Internet Governance' (2005) <http://www.wgig.org/docs/WGIGREPORT.pdf> accessed 31 August 2022

Books

Benedek W, and Kettemann M, *Freedom Of Expression And The Internet* (Council of Europe Publishing 2013)

Benesch S and others. *Reducing Online Hate Speech: Recommendations For Social Media Companies And Internet Intermediaries* (The Israel Democracy Institute 2020)

Bychawska-Siniarska D, *Protecting The Right To Freedom Of Expression Under The European Convention On Human Rights* (Council of Europe 2017)

George E, *Incorporating Rights* (5th edn, Oxford University Press 2021)

Guseva M and others, *Press Freedom And Development* (UNESCO 2008)

Hensley T, *The Boundaries Of Freedom Of Expression & Order In American Democracy* (Kent State University Press 2001)

Klug F, *The Three Pillars Of Liberty: Political Rights And Freedoms In The United Kingdom* (Routledge 1996)

Kurbalija J, *An Introduction To Internet Governance* (5th edn, DiploFoundation 2012)

Book Chapters

Bhagwat A, and Weinstein J, 'Freedom Of Expression And Democracy', *The Oxford Handbook of Freedom of Speech* (Oxford Academic 2021)

Braman S, 'Internet Policy', *The Handbook of Internet Studies* (Wiley-Blackwell 2010)

Journal Articles

Aceves W, 'Virtual Hatred: How Russia Tried To Start A Race War In The United States' (2019) 24 Michigan Journal of Race & Law

Albareda L, Lozano J, and Ysa T, 'Public Policies On Corporate Social Responsibility: The Role Of Governments In Europe' (2007) 74 Journal of Business Ethics

Alkiviadou N, 'Hate Speech On Social Media Networks: Towards A Regulatory Framework?' (2019) 28 Information & Communications Technology Law

Battaglini M, and Patacchini E, 'Social Networks In Policy Making' (2019) 11 Annual Review of Economics

Benkler Y, 'From Consumers To Users: Shifting The Deeper Structures Of Regulation Towards Sustainable Commons And User Access' (2000) 51 Federal Communications Law Journal

Comninos A, 'The Liability Of Internet Intermediaries In Nigeria, Kenya, South Africa And Uganda: An Uncertain Terrain' [2012] Intermediary Liability in Africa Research Papers

Davis R, 'The UN Guiding Principles On Business And Human Rights And Conflict-Affected Areas: State Obligations And Business Responsibilities' (2012) 94 International Review of the Red Cross

DeNardis L, 'The Emerging Field Of Internet Governance' [2010] Yale Information Society Project

Eisingerich A and others, 'Doing Good And Doing Better Despite Negative Information?: The Role Of Corporate Social Responsibility In Consumer Resistance To Negative Information' (2011) 14 Journal of Service Research

Elkins Z, and Ginsburg T, 'Imagining A World Without The Universal Declaration Of Human Rights' (2022) 74 World Politics

Giesler M, and Veresiu E, 'Creating The Responsible Consumer: Moralistic Governance Regimes And Consumer Subjectivity' (2014) 41 Journal of Consumer Research

Hannum H, 'The UDHR In National And International Law' (1998) 3 Health and Human Rights

Hefti A, and Ausserladscheider Jonas L, 'From Hate Speech To Incitement To Genocide: The Role Of The Media In The Rwandan Genocide' (2020) 38 Boston University International Law Journal

Heldt A, 'Reading Between The Lines And The Numbers: An Analysis Of The First Netzdg Reports' (2019) 8 Internet Policy Review

Kaesling K, 'Privatising Law Enforcement In Social Networks: A Comparative Model Analysis' (2018) 11 Erasmus Law Review

Kristin Timmermann W, 'Incitement In International Criminal Law' (2006) 88 International Review of the Red Cross

Martin Curran M, and Moran D, 'Impact of the Ftse4good Index on Firm Price: An Event Study' (2007) 82 Journal of Environmental Management

O'Laughlin B, 'Governing Capital? Corporate Social Responsibility And The Limits Of Regulation' (2008) 39 Development and Change

Palfrey J, 'Four Phases Of Internet Regulation' (2010) 77 Social Research: An International Quarterly

Sklar S, 'The Impact Of Social Media On The Legislative Process: How The Speech Or Debate Clause Could Be Interpreted' (2015) 10 Northwestern Journal of Law and Social Policy

Vargas Valdez J, 'Study On The Role Of Social Media And The Internet In Democratic Development' [2018] Venice Commission

Wilson D and others, 'Web 2.0: A Definition, Literature Review, And Directions For Future Research' [2011] AMCIS 2011 Proceedings - All Submissions

Ziewitz M, and Pentzold C, 'In Search Of Internet Governance: Performing Order In Digitally Networked Environments' (2014) 16 New Media & Society

Magazine and Newspaper Articles

Applebaum A, 'Ukraine And The Words That Lead To Mass Murder' *The Atlantic* (2022) <https://www.theatlantic.com/magazine/archive/2022/06/ukraine-mass-murder-hate-speech-soviet/629629/> accessed 31 August 2022

BBC, 'How Russian bots appear in your timeline' (2017) <https://www.bbc.com/news/technology-41982569> accessed 31 August 2022

Bump P, 'What Data On More Than 3,500 Russian Facebook Ads Reveals About The Interference Effort' *The Washington Post* (2018) <https://www.washingtonpost.com/news/politics/wp/2018/05/10/what-data-on-more-than-3500-russian-facebook-ads-reveals-about-the-interference-effort/> accessed 31 August 2022

Cieslak M, and Gerken T, 'Ukraine Crisis: Google Maps Live Traffic Data Turned Off In Country' *BBC* (2022) <https://www.bbc.com/news/technology-60561089> accessed 31 August 2022

Clayton J, 'How Kremlin Accounts Manipulate Twitter' *BBC* (2022) <https://www.bbc.com/news/technology-60790821> accessed 31 August 2022

Collins B, and Korecki N, 'Twitter Bans Over 100 Accounts That Pushed #IStandWithPutin' *CBS News* (2022) <https://www.nbcnews.com/tech/internet/twitter-bans-100-accounts-pushed-istandwithputin-rcna18655> accessed 31 August 2022

Culliford E, 'Twitter To Comply With EU Sanctions On Russian State Media' *Reuters* (2022) <https://www.reuters.com/technology/twitter-comply-with-eu-sanctions-russian-state-media-2022-03-02/> accessed 31 August 2022

Dave P, 'Google Blocks RT, Other Russian Channels From Earning Ad Dollars' *Reuters* (2022) <https://www.reuters.com/technology/youtube-blocks-rt-other-russian-channels-generating-revenue-2022-02-26/> accessed 31 August 2022

DW, 'Web Inventor Tim Berners-Lee Unveils Plan To Save The Internet' <https://www.dw.com/en/web-inventor-tim-berners-lee-unveils-plan-to-save-the-internet/a-51395985> accessed 31 August 2022

Ebbighausen R, 'Inciting Hatred Against Rohingya On Social Media' *DW* (2018) <https://www.dw.com/en/inciting-hatred-against-rohingya-on-social-media/a-45225962> accessed 31 August 2022

ERR, 'Estonia One Of Two EU Countries Not To Criminalize Hate Speech' (2020) <https://news.err.ee/1159938/estonia-one-of-two-eu-countries-not-to-criminalize-hate-speech> accessed 31 August 2022

Euronews, 'Pro-Kremlin Propaganda Still Rife On Tiktok Despite Ban' (2022) <https://www.euronews.com/my-europe/2022/04/13/pro-kremlin-propaganda-still-rife-on-tiktok-despite-ban-new-report-shows> accessed 31 August 2022

Financial Times, 'Holding Russia To Account For War Crimes' (2022) <https://www.ft.com/content/aacc2e1d-d450-4345-bedf-2cfb1af60e8a> accessed 31 August 2022

Folmar C, 'Multiple NGOs Including Amnesty International Forced To Shutter Offices In Russia' *The Hill* (2022) <https://thehill.com/policy/international/russia/3263143-multiple-ngos-including-amnesty-international-forced-to-shutter-offices-in-russia/> accessed 31 August 2022

Gilbert D, 'Russian Tiktok Influencers Are Being Paid To Spread Kremlin Propaganda' Vice (2022) <https://www.vice.com/en/article/epxken/russian-tiktok-influencers-paid-propaganda> accessed 31 August 2022

Graham T, and Thompson J, 'Russian Government Accounts Are Using A Twitter Loophole To Spread Disinformation' *The Conversation* (2022) <https://theconversation.com/russian-government-accounts-are-using-a-twitter-loophole-to-spread-disinformation-178001> accessed 31 August 2022

Grynbaum M, Koblin J, and Hsu T, 'Several Western News Organizations Suspend Operations In Russia' *The New York Times* (2022) <https://www.nytimes.com/2022/03/04/business/western-media-operations-russia.html> accessed 31 August 2022

Halpert M, 'Russia's Invasion Has Cost Ukraine Up To $600 Billion, Study Suggests' *Forbes* (2022) <https://www.forbes.com/sites/madelinehalpert/2022/05/04/russias-invasion-has-cost-ukraine-up-to-600-billion-study-suggests/> accessed 31 August 2022

Hofmann F, 'The Hybrid War That Began Before Russia Invaded Ukraine' *DW* (2022) <https://www.dw.com/en/hybrid-war-in-ukraine-began-before-russian-invasion/a-60914988> accessed 31 August 2022

Interfax-Ukraine, 'Ukraine's Dpty PM Asks Intl Services, Social Networks To Block Their Content In Russia' (2022) <https://en.interfax.com.ua/news/economic/802555.html> accessed 31 August 2022

Kottasova I, 'Russia Accused Of Inciting Genocide In Ukraine In New Report' *CNN* (2022) <https://www.ctvnews.ca/world/russia-accused-of-inciting-genocide-in-ukraine-in-new-report-1.5920894> accessed 31 August 2022

Luhn A, 'Ukraine blocks popular social networks as part of sanctions on Russia' *The Guardian* (2017)

Milmo D, 'Rohingya Sue Facebook For £150Bn Over Myanmar Genocide' *The Guardian* (2021) <https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence> accessed 31 August 2022

NOS Nieuws, 'Milieudefensie Dagvaardt Shell In Rechtszaak Om Uitstoot' (2019) <https://nos.nl/artikel/2279155-milieudefensie-dagvaardt-shell-in-rechtszaak-om-uitstoot> accessed 31 August 2022

Oremus W, 'Ukraine Says Big Tech Has Dropped The Ball On Russian Propaganda' *The Washington Post* (2022) <https://www.washingtonpost.com/technology/2022/07/14/ukraine-takedown-requests-russia-propaganda/> accessed 31 August 2022

Parker C, 'Russia Has Incited Genocide In Ukraine, Independent Experts Conclude' *The Washington Post* (2022) <https://www.washingtonpost.com/world/2022/05/27/genocide-ukraine-russia-analysis/> accessed 31 August 2022

Radio Liberty, 'Putin Signs 'Harsh' Law Allowing Long Prison Terms For 'False News' About Army' (2022) <https://www.rferl.org/a/russia-military-false-news/31737627.html> accessed 31 August 2022

Republicworld, 'UNSC To Consider 'Incitement To Violence' Among Reasons Behind Russia's War In Ukraine' (2022) <https://www.republicworld.com/world-news/russia-ukraine-crisis/unsc-to-consider-incitement-to-violence-among-reasons-behind-russias-war-in-ukraine-articleshow.html> accessed 31 August 2022

Sauer P, 'Russia Bans Facebook And Instagram Under 'Extremism' Law' *The Guardian* (2022) <https://www.theguardian.com/world/2022/mar/21/russia-bans-facebook-and-instagram-under-extremism-law> accessed 31 August 2022

Shepardson D, and Wang E, 'U.S. Senators Press Tiktok On Whether It Allows Russian "Pro-War Propaganda"' *Reuters* (2022) <https://www.reuters.com/business/media-telecom/us-senators-press-tiktok-whether-it-allows-russian-pro-war-propaganda-2022-06-17/> accessed 31 August 2022

The New Yorker, 'Copenhagen, Speech, And Violence' (2015) <https://www.newyorker.com/news/news-desk/copenhagen-speech-violence> accessed 31 August 2022

Tworek H, 'A Lesson From 1930S Germany: Beware State Control Of Social Media' *The Atlantic* (2019) <https://www.theatlantic.com/international/archive/2019/05/germany-war-radio-social-media/590149/> accessed 31 August 2022

Ukrinform, 'Russia's State-Orchestrated Incitement To Genocide Of Ukrainians' (2022) <https://www.ukrinform.net/rubric-ato/3495013-russias-stateorchestrated-incitement-to-genocide-of-ukrainians.html> accessed 31 August 2022

Wesolowski K, 'Fact Check: Fake News Thrives Amid Russia-Ukraine War' *DW* (2022) <https://www.dw.com/en/fact-check-fake-news-thrives-amid-russia-ukraine-war/a-61477502> accessed 31 August 2022

Blog Posts

Abbate J, 'The Internet: Global Evolution And Challenges' <https://www.bbvaopenmind.com/en/books/frontiers-of-knowledge/> accessed 31 August 2022.

Archyworldys, 'Telegram Should Adhere To The Netzdg' (2021) <https://www.archyworldys.com/telegram-should-adhere-to-the-netzdg/> accessed 31 August 2022

'Austria: The Draft Communication Platforms' Act Fails To Protect Freedom Of Expression' <https://www.article19.org/resources/austria-draft-communication-platforms-act-fails-freedom-of-expression/> accessed 31 August 2022

Benson S, 'Twitter To Label All State-Affiliated Russia Media' <https://www.politico.com/news/2022/02/28/twitter-label-state-affiliated-russia-media-00012351> accessed 31 August 2022

Brown D, 'Big Tech's Heavy Hand Around The Globe' <https://www.hrw.org/news/2020/09/08/big-techs-heavy-hand-around-globe> accessed 31 August 2022

'Censorship: Russia Blocks Access To Independent Media Over War Coverage' <https://europeanjournalists.org/blog/2022/03/01/censorhip-russia-blocks-access-to-independent-media-over-war-coverage/> accessed 31 August 2022

'Climate Change Actions Against Corporations: Milieudefensie Et Al. V. Royal Dutch Shell Plc.' <https://www.cliffordchance.com/insights/resources/blogs/business-and-human-

rights-insights/2021/01/climate-change-actions-against-corporations-milieudefensie-et-al-v-royal-dutch-shell-plc.html> accessed 31 August 2022

De León E, 'Russian Propaganda Is Thriving In Spanish' <https://slate.com/technology/2022/06/kremlin-propaganda-spanish-latam.html> accessed 31 August 2022

Desai S, 'PACE Calls For International Tribunal To Probe Russian War Crimes In Ukraine' <https://www.aa.com.tr/en/europe/pace-calls-for-international-tribunal-to-probe-russian-war-crimes-in-ukraine/2575904 (Last accessed: 25 August 2022)> accessed 31 August 2022

Doyle K, 'BHR In The Tech Sector: Much To Celebrate, More To Do' <https://www.gp-digital.org/bhr-in-the-tech-sector-much-to-celebrate-more-to-do/> accessed 31 August 2022

'France: Constitutional Council Declares French Hate Speech 'Avia' Law Unconstitutional' <https://www.article19.org/resources/france-constitutional-council-declares-french-hate-speech-avialaw-unconstitutional/> accessed 31 August 2022

Heller K, 'Creating A Special Tribunal For Aggression Against Ukraine Is A Bad Idea' <https://opiniojuris.org/2022/03/07/creating-a-special-tribunal-for-aggression-against-ukraine-is-a-bad-idea/ (Last accessed: 25 August 2022)> accessed 31 August 2022

Johnson A, and Castro D, 'How Other Countries Have Dealt With Intermediary Liability' *Information Technology & Innovation Foundation* (2021) <https://itif.org/publications/2021/02/22/how-other-countries-have-dealt-intermediary-liability/> accessed 31 August 2022

Jones S, 'The Social Dilemma And The Human Rights Risks Of Big Tech' <https://www.humanrightspulse.com/mastercontentblog/the-social-dilemma-and-the-human-rights-risks-of-big-tech> accessed 31 August 2022

Karanicolas M, 'Newly Published Citizens Protection (Against Online Harm) Rules Are A Disaster For Freedom Of Expression In Pakistan' <https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-information/wiii-blog/newly-published-citizens-protection-against-online-harm-rules-are-disaster-freedom-expression> accessed 31 August 2022

Makridin I, 'Banned" Russian Propaganda Still Easy To Find On YouTube' <https://www.codastory.com/newsletters/russian-propaganda-youtube/> accessed 31 August 2022

Novak M, 'YouTube Stops Monetization For Video Creators In Russia But There's One Exception' <https://gizmodo.com/youtube-stops-monetization-for-video-creators-in-russia-1848633335> accessed 31 August 2022

O'Callaghan S, 'Can The EU Digital Services Act Succeed In Controlling Big Tech And Protecting Consumer Rights?' <https://www.iglobenews.org/can-the-eu-digital-services-act-succeed-in-controlling-big-tech-and-protecting-consumer-rights/> accessed 31 August 2022

'Russia Criminalizes Independent War Reporting, Anti-War Protests' <https://www.hrw.org/news/2022/03/07/russia-criminalizes-independent-war-reporting-anti-war-protests> accessed 31 August 2022

'Russia's Crackdown On Independent Media And Access To Information Online' <https://www.csis.org/analysis/russias-crackdown-independent-media-and-access-information-online> accessed 31 August 2022

Scott M, 'Telegram Bans Russian State Media After Pressure From Europe' <https://www.politico.eu/article/russia-rt-media-telegram-ukraine/> accessed 31 August 2022

Staber G, 'Communication Platforms Face New Obligations And High Fines In Austria' <https://www.lexology.com/library/detail.aspx?g=fcf46df4-4694-4f10-b11b-67564a824470> accessed 31 August 2022

'Tech Giants And Human Rights: Investor Expectations' <https://www.humanrights.dk/sites/humanrights.dk/files/media/document/Tech%20giants%20and%20human%20rights_2021.pdf> accessed 31 August 2022

'Ukraine: Briefing On 'Incitement To Violence Leading To Atrocity Crimes' <https://reliefweb.int/report/ukraine/ukraine-briefing-incitement-violence-leading-atrocity-crimes> accessed 31 August 2022

'Updates: Digital Rights In The Russia-Ukraine Conflict' <https://www.accessnow.org/digital-rights-ukraine-russia-conflict/> accessed 31 August 2022

Vile J, 'Incitement To Imminent Lawless Action' <https://www.mtsu.edu/first-amendment/article/970/incitement-to-imminent-lawless-action> accessed 31 August 2022

Walker K, 'Helping Ukraine' <https://blog.google/inside-google/company-announcements/helping-ukraine/?utm_source=tw&utm_medium=social&utm_campaign=og&utm_content=&utm_term> accessed 31 August 2022

Website Content

(*Facebook*) <https://www.facebook.com/groups/mid.dnr> accessed 31 August 2022

(*Facebook*) <https://www.facebook.com/photo.php?fbid=117088227751551&set=pb.100083511233390.-2207520000..&type=3> accessed 31 August 2022

(*Facebook*) <https://www.facebook.com/RusovDvizhenie/photos/145264214782351> accessed 31 August 2022

(*Facebook*) <https://www.facebook.com/RusovDvizhenie/posts/pfbid02MdwgVgHhBjMyCBmnV3ujx17AWBB1v8asGWmvb9FZTaHRz6ErXvGSmGF49k8SKPc5l> accessed 31 August 2022

(*Facebook*) <https://www.facebook.com/RusovDvizhenie> accessed 31 August 2022

(*Facebook*) <https://www.facebook.com/RussianEmbassy> accessed 31 August 2022

(*Instagram*) <https://www.instagram.com/p/Cek-gMUsBMJ/> accessed 31 August 2022

(*Instagram*) <https://www.instagram.com/p/CeW3nk3jsJP/> accessed 31 August 2022

(*Instagram*) <https://www.instagram.com/p/ChfQseBoEGE/> accessed 31 August 2022

(*Telegram*) <https://t.me/deathhohlam> accessed 31 August 2022

(*Telegram*) <https://t.me/Hohli_pidorasi> accessed 31 August 2022

(*Telegram*) <https://t.me/hohlopidorislavarossii> accessed 31 August 2022

(*Telegram*) <https://t.me/margaritasimonyan> accessed 31 August 2022

(*Telegram*) <https://t.me/otborsalo> accessed 31 August 2022

(*Telegram*) <https://t.me/trupvsy1> accessed 31 August 2022

(*Telegram*) <https://t.me/ukropi> accessed 31 August 2022

(*Telegram*) <https://t.me/vladlentatarsky> accessed 31 August 2022

(*Telegram*) <https://t.me/vysokygovorit> accessed 31 August 2022

(*TikTok*) <https://www.tiktok.com/@russian_military_guy> accessed 31 August 2022

(*TikTok*) <https://www.tiktok.com/@russian_mma> accessed 31 August 2022

(*Twitter*) <https://twitter.com/BobbyAllyn/status/1498421323135012865> accessed 31 August 2022

(*Twitter*) <https://twitter.com/M_Simonyan> accessed 31 August 2022

(*Twitter*) <https://twitter.com/MirelaBjelic/status/486245763669053440> accessed 31 August 2022

(*Twitter*) <https://twitter.com/ngleicher/status/1497417241947607043> accessed 31 August 2022

(*Twitter*) <https://twitter.com/nickclegg/status/1498395147536527360> accessed 31 August 2022.

(*Twitter*) <https://twitter.com/RussianEmbassy> accessed 31 August 2022

(*Twitter*) <https://twitter.com/TwitterSafety/status/1497353968565075968?s=20&t=VwP5VUPj2QuDCgZhN9az2w> accessed 31 August 2022

(*Twitter*) <https://twitter.com/VladPunished/status/1561123899403894784> accessed 31 August 2022

(*Twitter*) <https://twitter.com/YouTubeInsider/status/1498772480034365440> accessed 31 August 2022

(*Twitter*) <https://twitter.com/YouTubeInsider/status/1498772481309437952?s=20&t=SQJVgNzHU6C1Hn6PGUI8rg> accessed 31 August 2022

(*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335030168899595> accessed 31 August 2022

(*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335085122666500> accessed 31 August 2022

(*Twitter*) <https://twitter.com/YouTubeInsider/status/1502335119914381314?s=20&t=j_NJ6tVDMG4MQOPjRSxelA> accessed 31 August 2022

'About' (*Covalence*) <https://www.covalence.ch/index.php/about-us/> accessed 31 August 2022

'About GNI' (*Global Networks Initiative*) <https://globalnetworkinitiative.org/about-gni/> accessed 31 August 2022

'About Government And State-Affiliated Media Account Labels On Twitter' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/state-affiliated> accessed 31 August 2022

'Acerca De LACNIC' (*LACNIC*) <https://www.lacnic.net/966/1/lacnic/acerca-de-lacnic> accessed 31 August 2022

'AFRINIC-31' (*AFRINIC*) <https://meeting.afrinic.net/afrinic-31/en/about/afrinic-31> accessed 31 August 2022

'Alert (AA22-110A)' (*Cybersecurity & Infrastracture Security Agency*) <https://www.cisa.gov/uscert/ncas/alerts/aa22-110a> accessed 31 August 2022

'Anatolii Sharii's YouTube Channel' (*YouTube*) <https://www.youtube.com/user/SuperSharij> accessed 31 August 2022

'APNIC Serves The Asia Pacific Region' (*APNIC*) <https://www.apnic.net/about-apnic/organization/apnic-region/> accessed 31 August 2022

'Austria Communication Platform Act' (*Facebook*) <https://www.facebook.com/help/3846278558774584/?helpref=hc_fnav> accessed 31 August 2022

'Bringing more context to content on TikTok' (*TikTok*) <https://newsroom.tiktok.com/en-us/bringing-more-context-to-content-on-tiktok?utm_source=COMMSTWITTER&utm_medium=social&utm_campaign=030622> accessed 31 August 2022

'Commission Staff Working Document - Corporate Social Responsibility, Responsible Business Conduct, And Business And Human Rights: Overview Of Progress' (*European Commission*, 2019) <https://ec.europa.eu/docsroom/documents/34482> accessed 31 August 2022

'Community Guidelines' (*Instagram*) <https://help.instagram.com/477434105621119/?helpref=hc_fnav> accessed 31 August 2022

'Community Guidelines' (*TikTok*) <https://www.tiktok.com/community-guidelines#29> accessed 31 August 2022

'Community Guidelines' (*YouTube*) <https://www.youtube.com/intl/en_us/howyoutubeworks/policies/community-guidelines/#community-guidelines> accessed 31 August 2022

'Conventions And Recommendations' (*International Labor Organization*) <https://www.ilo.org/global/standards/introduction-to-international-labour-standards/conventions-and-recommendations/lang--en/index.htm> accessed 31 August 2022

'Corporate Documents' (*ARIN*) <https://www.arin.net/about/corporate/documents/> accessed 31 August 2022

'Corporate Human Rights Policy' (*Meta*) <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf> accessed 31 August 2022

'Corporate Social Responsibility & Responsible Business Conduct' (*European Commission*) <https://single-market-economy.ec.europa.eu/industry/sustainability/corporate-social-responsibility-responsible-business-conduct_en> accessed 31 August 2022

'Countries' (*National Action Plans on Business and Human Rights*) <https://globalnaps.org/country/> accessed 31 August 2022

'Creating A Common Good Balance Sheet' (*Economy for the Common Good*) <https://web.archive.org/web/20130426095936/http:/economia-del-bene-comune.it/en/content/creating-common-good-balance-sheet> accessed 31 August 2022

'Defending And Respecting The Rights Of People Using Our Service' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice> accessed 31 August 2022

'Digital Services: Landmark Rules Adopted For A Safer, Open Online Environment' (*European Parliament*, 2022) <https://www.europarl.europa.eu/news/en/press-room/20220701IPR34364/digital-services-landmark-rules-adopted-for-a-safer-open-online-environment> accessed 31 August 2022

'Digital Vs Traditional Media Consumption' (*globalwebindex*) <https://www.amic.media/media/files/file_352_2142.pdf> accessed 31 August 2022

'Facebook Community Standards' (*Facebook*) <https://transparency.fb.com/policies/community-standards/> accessed 31 August 2022

'FAQS' (*Corporate Register*) <https://www.corporateregister.com/about/> accessed 31 August 2022

'Featured Policies – Hate Speech' (*Google Transparency Report*) <https://transparencyreport.google.com/youtube-policy/featured-policies/hate-speech?hl=en> accessed 31 August 2022

'Freedom Of Expression - Article 10' (*Council of Europe*) <https://www.coe.int/en/web/human-rights-convention/expression> accessed 31 August 2022

'Freedom Of Expression On The Internet' (*UNESCO*) <https://en.unesco.org/themes/freedom-expression-internet> accessed 31 August 2022

'Hate Speech – Data' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/#data> accessed 31 August 2022

'Hate Speech' (*Meta*) <https://transparency.fb.com/de-de/policies/community-standards/hate-speech/> accessed 31 August 2022

'Hate Speech Against Ukrainians In Social Networks' (*Google Drive*, 2022) <https://drive.google.com/drive/folders/1qe1mtQkJqMi8QLe9JTxLeK2i34a9aAD5?usp=sharing> accessed 31 August 2022

'Hate Speech Is Rising Around The World' (*UN*) <https://www.un.org/en/hate-speech> accessed 31 August 2022

'Hate Speech Policy' (YouTube Help) <https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436> accessed 31 August 2022

'Hateful Behavior' (*TikTok*) <https://www.tiktok.com/community-guidelines#38> accessed 31 August 2022

'Hateful Conduct Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy> accessed 31 August 2022

'How Many People Use Telegram In 2022? 55 Telegram Stats' (*Backlinko*) <https://backlinko.com/telegram-users> accessed 31 August 2022

'How to Report Things' (*Facebook*) <https://www.facebook.com/help/reportlinks/> accessed 31 August 2022

'Human Rights' (*Google*) <https://about.google/human-rights/> accessed 31 August 2022

'Human Rights Law' (*UN*) <https://www.un.org/ruleoflaw/thematic-areas/international-law-courts-tribunals/human-rights-law/> accessed 31 August 2022.

'IGF 2021 WS #57 Multistakeholder Initiatives In Content Governance' (*IGF Internet Governance Forum*) <https://www.intgovforum.org/multilingual/content/igf-2021-ws-57-multistakeholder-initiatives-in-content-governance> accessed 31 August 2022

'Index Inclusion Rules For The Ftse4good Index Series V2.0' (*FTSE Russell*) <https://web.archive.org/web/20171215102622/http:/www.ftse.com/products/downloads/F4G-Index-Inclusion-Rules.pdf> accessed 31 August 2022

'ISIS Watch' (*Telegram*) <https://t.me/isiswatch> accessed 31 August 2022

Kenton W, 'Web 2.0' (*Investopedia*) <https://www.investopedia.com/terms/w/web-20.asp> accessed 31 August 2022

'Legal Removal Request' *(Instagram)* <https://help.instagram.com/874680996209917/?helpref=hc_fnav> accessed 31 August 2022

'Letter To Social Media Platforms On Crisis Zones' (*Electronic Frontier Foundation*, 2022) <https://www.eff.org/document/letter-social-media-platforms-crisis-zones> accessed 31 August 2022

'Mandate Of The Commission' (African Commission on Human and Peoples' Rights) <https://www.achpr.org/mandateofthecommission> accessed 31 August 2022

'Mark Zuckerberg Stands For Voice And Free Expression' (*Meta*, 2019) <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/> accessed 31 August 2022

'Meta's Ongoing Efforts Regarding Russia's Invasion Of Ukraine' (*Meta*, 2022) <https://about.fb.com/news/2022/02/metas-ongoing-efforts-regarding-russias-invasion-of-ukraine/> accessed 31 August 2022

'NGOs Call On Social Media Platforms To Strengthen Human Rights Due Diligence In Crisis Situations; Incl. Co. Response' (*Business and Human Rights Resource Center*, 2022) <https://www.business-humanrights.org/en/latest-news/ngos-call-on-social-media-platforms-to-strengthen-their-human-rights-due-diligence-and-address-structural-inequalities-in-conflict-zones/> accessed 31 August 2022

'Nuremberg Trial Proceedings Vol. 12' (*The Avalon Project*) <https://avalon.law.yale.edu/imt/04-29-46.asp> accessed 31 August 2022

Obermair D, 'Tim Berners-Lee: The Goal Of The Worldwide Web Is To Serve Humanity' <https://www.instituteofnext.com/tim-berners-lee-the-goal-of-the-worldwide-web-is-to-serve-humanity/> accessed 31 August 2022

'Other Legal Complaint' (*YouTube Help*) <https://support.google.com/youtube/contact/other_legal> accessed 31 August 2022

'Our Commitment To Human Rights' (*Apple*) <https://s2.q4cdn.com/470004039/files/doc_downloads/gov_docs/Apple-Human-Rights-Policy.pdf> accessed 31 August 2022

'Penetration Of Selected Social Media Platforms In Ukraine And Russia As Of November 2021' (*Statista*) <https://www.statista.com/statistics/1308258/social-media-penetration-ukraine-russia/> accessed 31 August 2022

'Removals Under The Network Enforcement Law' (*YouTube Transparency Report*) <https://transparencyreport.google.com/netzdg/youtube?hl=us> accessed 31 August 2022

'Report a problem' (*TikTok*) <https://support.tiktok.com/en/safety-hc/report-a-problem> accessed 31 August 2022

'Report content under the Network Enforcement Law' (*YouTube Help)* <https://support.google.com/youtube/contact/netzdg> accessed 31 August 2022

'Report Inappropriate Videos, Channels, And Other Content On YouTube' (*YouTube Help*) <https://support.google.com/youtube/answer/2802027?hl=en&co=GENIE.Platform%3DDesktop&oco=0#zippy=%2Cprivacy-reporting%2Clegal-reporting%2Creport-a-video%2Creport-a-channel%2Creport-a-playlist%2Creport-a-thumbnail%2Creport-a-comment%2Creport-a-link%2Creport-a-live-chat-message%2Creport-an-ad (Last accessed: 25 August 2022)> accessed 31 August 2022

'Report Something' (*Facebook*) <https://www.facebook.com/help/263149623790594?ref=tc> accessed 31 August 2022

'Response From Twitter To Letter Calling On Social Media Platforms To Strengthen Their Human Rights Due Diligence' (*Business and Human Rights Resource Center*, 2022) <https://www.business-humanrights.org/en/latest-news/response-from-twitter-to-letter-calling-social-media-platforms-for-long-term-investment-in-human-rights/> accessed 31 August 2022

'Share your feedback' (*TikTok*) <https://www.tiktok.com/legal/report/feedback> accessed 31 August 2022

'Staying Safe On Twitter And Sensitive Content' (*Twitter*) <https://help.twitter.com/en/forms/safety-and-sensitive-content/abuse/legal-rep> accessed 31 August 2022

'Support Inbox' (*Facebook*) <https://www.facebook.com/support> accessed 31 August 2022

'Telegram FAQ' (*Telegram*) <https://telegram.org/faq> accessed 31 August 2022

'Terms And Policies' (*Google*) <https://support.google.com/googlecurrents/answer/9680387?hl=en> accessed 31 August 2022

'Text-Based Legal Complaints Under Kopl-G' (*YouTube Help*) <https://support.google.com/youtube/answer/11052657?hl=en&ref_topic=6154211> accessed 31 August 2022

'The Digital Markets Act: Ensuring Fair And Open Digital Markets' (*European Commission*) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en> accessed 31 August 2022

'The Digital Services Act: Ensuring A Safe And Accountable Online Environment' (*European Commission*) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en> accessed 31 August 2022

'The Foundation Of International Human Rights Law' (UN) <https://www.un.org/en/about-us/udhr/foundation-of-international-human-rights-law> accessed 31 August 2022

'The Role Of The High Commissioner For Human Rights' (*OHCHR*) <https://www.ohchr.org/en/about-us/high-commissioner> accessed 31 August 2022

'Ukraine: Situation In Ukraine - ICC-01/22' (*International Criminal Court*, 2022) <https://www.icc-cpi.int/ukraine> accessed 31 August 2022

'United Nations Strategy And Plan Of Action On Hate Speech' (*UN*, 2019) <https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf> accessed 31 August 2022

'Updates on Our Security Work in Ukraine' (*Meta*, 2022) <https://about.fb.com/news/2022/02/security-updates-ukraine/> accessed 31 August 2022

'Violence And Incitement' (*Meta*) <https://transparency.fb.com/policies/community-standards/violence-incitement/#policy-details> accessed 31 August 2022

'Violent Extremism' (*TikTok*) <https://www.tiktok.com/community-guidelines#39> accessed 31 August 2022

'Violent Threats Policy' (*Twitter*) <https://help.twitter.com/en/rules-and-policies/violent-threats-glorification> accessed 31 August 2022

'We Are Committed To Protecting Your Voice And Helping You Connect And Share Safely' (*Meta*) <https://about.facebook.com/actions/promoting-safety-and-expression/> accessed 31 August 2022

'What We Do' (ASEAN) <https://asean.org/what-we-do/> accessed 31 August 2022

'What We Do' (OAS) <https://www.oas.org/en/about/what_we_do.asp> accessed 31 August 2022

'What We Do' (*RIPE NCC*) <https://www.ripe.net/about-us/what-we-do> accessed 31 August 2022

'What We Do' (UNICEF) <https://www.unicef.org/what-we-do> accessed 31 August 2022

'What's The Difference Between CPA And The Instagram Community Guidelines?' (*Instagram*) <https://help.instagram.com/428536715033518/?helpref=related_articles> accessed 31 August 2022

'What's The Difference Between NetzDG And The Instagram Community Guidelines?' (*Instagram*) <https://help.instagram.com/1787585044668150/?helpref=related_articles> accessed 31 August 2022.

'Who We Are' (*IETF*) <https://www.ietf.org/about/who/> accessed 31 August 2022

'YouTube Community Guidelines Enforcement' (*Google Transparency Report*) <https://transparencyreport.google.com/youtube-policy/removals> accessed 31 August 2022