The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism

Analysis on Certain Central European Countries

Studies of the Central European Professors' Network
ISSN 2786-2518

Description
The book series *Studies of the Central European Professors' Network* publishes the results of research by members of the Central European Professors' Network with the coordination of the Budapest-based Ferenc Mádl Institute of Comparative Law. The primary aim of the series is to present and address legal issues that are strongly related to the Central European region, taking into account the particular legal traditions, culture, and approach of the countries therein. The authenticity of the books can be seen in the fact that renowned authors from the Central European region write about the legal instruments of countries of the Central European region in English. The book series aims to establish itself as a comparative legal research forum by contributing to the stronger cooperation of the countries concerned and by ensuring the "best practices" and making different legal solutions available and interpretable to all of the states in Central Europe. However, it also aims to provide insights and detailed analyses of these topics to all interested legal scholars and legal practitioners outside the region so that they might become acquainted with the legal systems of Central European countries regarding a great variety of subjects.

The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism

Analysis on Certain Central European Countries

Edited by
Marcin Wielec

FERENC MÁDL
INSTITUTE OF COMPARATIVE LAW

CEA
PUBLISHING

BUDAPEST – MISKOLC | 2021

STUDIES OF THE CENTRAL EUROPEAN
PROFESSORS' NETWORK

The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism
Analysis on Certain Central European Countries

Published by

© Ferenc Mádl Institute of Comparative Law
(Budapest, Hungary)
ISBN 978-615-6356-08-6
ISBN 978-615-6356-09-3 (eBook)

and

Central European Academic Publishing
(Miskolc, Hungary)
ISBN 978-615-01-3007-1
ISBN 978-615-01-3008-8 (eBook)

DOI: 10.54237/profnet.2021.mwsm

The address of Ferenc Mádl Institute of Comparative Law: 1123 Budapest, Alkotás str. 55 (Hungary)
The address of Central European Academic Publishing: 3515 Miskolc-Egyetemváros, Building A/6 (Hungary)

The book was reviewed by
Anna-Maria Getoš Kalac (Associate Professor, University of Zagreb),
Radomír Jakab (Associate Professor, Pavol Jozef Šafárik University in Košice),
Slobodan Marković (Professor, University of Belgrade),
Anita Paulovics (Professor, University of Miskolc),
Marcin Rau (Assistant Professor, Cardinal Stefan Wyszynski University),
Květoslav Růžička (Professor, Charles University in Prague and Masaryk University),
Aleš Završnik (Professor, University of Ljubljana).

Contents

Notes on the Contributors

Editor

Marcin Wielec is Director of the Institute of Justice in Warsaw and Head of the Department of Criminal Procedure at the Faculty of Law and Administration of the Cardinal Stefan Wyszynski University in Warsaw. He is the author and co-author of several books and a number of scientific articles in the field of criminal proceedings. He is a member of the Scientific Council of the journal titled *Probacja*, a graduate of IESE Business School and KSAP. He is the author of expert opinions and legal opinions for public authorities. His general areas of research interest include criminal law and proceedings, disciplinary law and proceedings, executive criminal law, and in particular the areas of correlation of criminal proceedings with other areas of law such as canon law, administrative law, and disciplinary law, and issues related to the axiology of criminal proceedings.

Contributors

Kristina Čufar is Researcher at the Faculty of Law of the University of Ljubljana, where she teaches Introduction to Jurisprudence and Sociology of Law. She holds a Ph.D. from the European University Institute and was awarded grants to spend time as a visiting researcher at the London School of Economics and the Max Planck Institute for Comparative Public Law and International Law. Her research interests include law and technology, legal philosophy and theory, especially critical legal theory, and law and the humanities.

Davor Derenčinović is Professor of criminal law and criminology at the Faculty of Law of the University of Zagreb. He is the president of the Croatian Academy of Legal Science. He is an international expert and consultant for international organizations (United Nations, Council of Europe) and law firms for criminal law, human rights, and anticorruption. He was elected judge to the European Court of Human Rights in respect of Croatia for a term of office of nine years which shall commence on 2 January 2022. He is the president and member of the Council of Europe Group of Experts on Action against Trafficking in Human Beings (2009–2012, 2017–2020).

Gábor Hulkó is Associate Professor at and Deputy-Head of the Department of Administrative and Financial Law of the Faculty of Law of the Széchenyi István University in Győr and researcher at the Department of International and European Law of the Budapest-based Ferenc Mádl Institute of Comparative Law. His research interests cover media regulation, local government finances and regulation, as well as comparative methods in law. He is the author of several book chapters and more than 50 reviewed articles. He is Head of the Central and Eastern European Association of Public Administration and a member of the Information and Organization Centre for the Research on the Public Finances and Tax Law in the Countries of Central and Eastern Europe and the Observatory on Local Autonomy researchers network. He is editor-in-chief of the scholarly journal *Institutiones Administrationis – Journal of Administrative Sciences*, and member of several other editorial boards of scholarly journals, such as *In Medias Res*, *KözigazgatásTudomány*, *Central European Journal of Comparative Law*, *Procedury administracyjne i podatkowe* and *Annual Center Review*.

András Koltay is Rector and Professor of the National University of Public Service (Budapest). He has been a lecturer at the Faculty of Law of the Pázmány Péter Catholic University in Budapest since 2002. In 2018, he was appointed professor of law. He received his LL.M. degree in public law at the University College London in 2006, and PhD degree in law at the Pázmány Péter Catholic University in 2008. He attended the Human Rights Course of the International Institute of Human Rights in Strasbourg in 2003. His principal research has concerned the freedom of speech, personality rights, and media regulations, but he also deals with other constitutional questions. He is the author of more than 350 publications and numerous monographs on freedom of speech; in English: *Freedom of Speech: The Unreachable Mirage* (Wolters Kluwer 2013), *The Troubled Relationship between Religions and the State: Freedom of Expression and Freedom of Religion* (Whitelocke 2017), and *New Media and Freedom of Expression* (Hart 2019). He has been a speaker at more than 125 conferences in several countries.

Bartłomiej Oręziak is Coordinator of the Institute of Justice in Warsaw's Center for Strategic Analysis, laureate of the Minister of Science and Higher Education Scholarship for outstanding achievements in science for the 2017/2018 academic year, winner of the DOCUP 2020 competition, member of the Artificial Intelligence Working Group at the Chancellery of the President Council of Ministers, author of several dozen scientific texts (articles, chapters, glosses, etc.), head or member of international and national research projects, and a speaker and organizer of international and national scientific conferences. His main interests cover the relationship between new technologies and law, human rights protection, intellectual property law, public international law, and criminal law and trial.

Dušan V. Popović is Professor of intellectual property law, competition law, and internet law at the Faculty of Law of the University of Belgrade. He holds a PhD degree from the University Paris-Nanterre, LL.M. degree from the University of Nancy, and LL.B. degree from the University of Belgrade. He was a visiting researcher at the University Panthéon-Assas (2019), Max Planck Institute for Innovation and Competition (2014, 2010), CEIPI—University of Strasbourg (2012, 2011, 2010), and University of Salzburg (2008). He was a visiting professor at the University of Lyon III Jean Moulin (2018/2019) and University of Skopje (2014–2016), and held a number of guest lectures at different European universities. He served as a Jean Monnet Module Leader within the Erasmus+ project Free Trade Agreements and European Integration of SEE Countries (2017–2020).

Aleš Rozehnal is Associate Professor of commercial law at the Faculty of Law of the Charles University. He is the author of several publications, such as *Loans Secured by Lien*, *Corruption the Czech Way*, *Media Law*, *Media Laws—Commentary*, *Media Law in the Czech Republic*, *Commentary on the Business Corporations Act*, *Substantive Civil Law*, *Law of Obligations*, *Commercial Law*, *Civil Procedure Strategy*, and *Compensation of Reflective Loss Incurred by Corporate Partners*. He works as an external employee of the Faculty of Law of Charles University. He worked for seven years as an assistant professor at the Faculty of Law of the University of West Bohemia in Pilsen.

Sanja Savčić is Associate Professor at the Faculty of Law of the University of Novi Sad. After graduation, she clerked at the Commercial Court in Novi Sad. She received her LL.M. degree from the Faculty of Law of the University of Novi Sad with a thesis titled *Compensation for Material Damages in Copyright Law*. She conducted research at the Max Planck Institute for Intellectual Property in Munich from October 2004 to February 2005 and from June to October 2005. In 2010, at the Faculty of Law of the University of Belgrade, she defended her doctoral dissertation '*Software License Agreement*', for which her degree was awarded with distinction. Her areas of academic interest include general theory of the law of obligations, contract law, and intellectual property.

Chapter I

Fake News as an Important Factor in Digital Platforms' and Social Media's Impact on the Guarantees of Freedom of Expression and the Truth of Information

Marcin Wielec

1. Review of digital platforms and social media in Poland

1.1 Digitization and globalization as a domain of emergence and operation of digital platforms

Nowadays, people have no choice but to use technological advances. In contrast to five or ten years ago, the present is a completely different reality in terms of technological possibilities, the dissemination of information, forging personal relationships, networking, etc. The fast-paced emergence of technological innovations means that everyday and professional life in communities has—or, it would seem, should—become, easier, more interesting, and above all, more effective.

The goal of new technologies, broadly understood, is precisely to shorten certain social and professional distances and facilitate processes that, so far, have been indispensable but highly complex. It is rightly assumed that "new information technologies, with the Internet at the forefront, spread around the world in less than 20

Marcin Wielec (2021) Fake News as an Important Factor in Digital Platforms' and Social Media's Impact on the Guarantees of Freedom of Expression and the Truth of Information. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 9–46. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_1

years, from the mid-1970s to the mid-1990s"[1] and have become usual amenities used in everyday life.

One of the outcomes brought about by these changes is the emergence of a parallel, and sometimes even alternative, reality. This is linked to the emergence, rapid growth, and deployment of digitization and digitalization, which have entered various areas of life with extraordinary impetus, bringing solutions that were previously deemed impossible to design and implement. These new possibilities have led to the emergence of a different reality that offers previously unknown options for designing and performing ordinary activities.

For the purposes of this analysis, the fundamental question is what digitization and digitalization actually are, as these appear to be the driving force and the basis for change. They provide the inherent ecosystem (zone) in which momentous changes have and will continue to take place. The zone delimited by digitization and digitalization has been occupied and leveraged, among others, by digital platforms, which have become one of the key accessories to the phenomenon of fake news that we set out to analyze in this paper.

At first impression, when describing digitization as a mechanical activation of a series of tasks undertaken in succession, one may simply point out that it is actually a process inextricably linked with the transformation of the original form of some material or immaterial entity into a complex and new type of digital recording, in which the natural form of the entity being recorded is transformed into a numerical representation, that is, a specific and systemically ordered sequence of numerical values. For example, definitions have been put forward that "scans of historical documents published on the Internet are numbers in digital format, which, in order to be human readable, must be reconstructed using appropriate software."[2] This is the simplest possible illustration of the digitization process, which involves changing something's original form into a digital form.

This is the basic definition of the process, as it is assumes that digitization is "the transformation of any analogue form of a document (book, image, sound) into a binary form, (...) or rather digitization is equivalent to scanning analogue material and processing it into a digital form."[3] Going further into another domain, digitalization is the process of transforming individual analogue information streams into digital form, or the way in which countries, organizations, and companies adopt or increase the use of information and communication technologies (ICT).[4]

In dictionary terms, the Polish definition covers both digitization and digitalization, as *firstly;* a change of the form of a signal from analogue to digital, in the process of analogue-to-digital conversion; *secondly*; a set of activities aimed at replacing devices based on analogue technology in technical systems with digital

1 Warzecha, 2017, p. 85.
2 Wilkowski, 2013, p. 10.
3 Mejor, 2012, p. 265.
4 Kuźniar, 2019, p. 275.

systems (digital technology).[5] The concepts of digitization and digitalization are, therefore, semantically identical, so to avoid unnecessary complications in our analyses, these terms will be used interchangeably.

Without delving into the definitional aspects of digitization or digitalization, one must note that this process has undoubtedly activated a range of other processes in various areas of life that are often incompatible with each other. It is the emerging technological opportunities leading directly to the initiation and dynamic growth of the digitization process that make it the first element in the complicated structure of the creation and operation of digital platforms, which we will, in this analysis, identify as the carriers or vehicles of the fake news phenomenon.

Digitization has therefore become a fact of life, but one must remember that in the context of fake news, there is one more component that is at play with the digitization process, that is, the process of globalization.

The term 'globalization' also encompasses different conceptual meanings. According to the simplest of definitions, it can be best described as a process through which the world is increasingly becoming a single place.[6] In dictionary terms, Polish studies point out that globalization is defined as a process involving, *inter alia*, an increase in the turnover of international trade, the flows of capital, and people and technology, as well as the blurring of cultural differences.[7] Globalization assumes the standardization of specific activities aimed at achieving the planned effects on the largest possible scale and is associated—as demonstrated by the attempts to define it in dictionaries—first with commercial and economic processes, then with social and cultural processes, and finally, with technological processes. The globalization phenomenon leads to standardization through the internationalization of, for example, commercial activities, where the visible effect is the presence of the same business entities operating in the broadest possible market and participating in information activities, entertainment or technology activities, etc. Consequently, the ongoing and never-ending process of globalization gives societies in different countries options for purchasing goods and using various other services or flows of people or information to exactly the same extent. This saves a lot of time and effort and makes space barriers obsolete. It is argued that globalization as primarily associated with international trade relations is characterized by the following:

> *first*; increasing mobility of capital and goods, and even services, treated before as non-commercial; *second*; it is accompanied by technical progress on an unprecedented scale, especially the rapid spread of innovation, *third*; sharply reduced transaction costs of economic cooperation with foreign countries, including above all the cost of transport and communication.[8]

5 See: https://encyklopedia.pwn.pl/haslo/cyfryzacja;4007905.html.
6 Kaczmarek, 2014, p. 35.
7 See: http://sjp.pwn.pl/sjp/globalizacja;2559335.
8 Czarny, 2014, p. 5.

With the two components seen through the prism of digitization and globalization, one can see globalization as the bedrock of digitization, given that instruments are needed to bring to effect its fundamental assumption of the standardization of certain relations, activities, and objectives. Each stage of globalization has stimulated research and inventions for making it fast and effective. However, in this maze of various globalization components, the transfer of information, its subsequent appropriate interpretation, and its global dissemination have always been key.

Therefore, what comes to mind is an interplay of components that comprise the foundation for the creation and action of fake news, which is one of the links in this ecosystem.

The interacting components are as follows:

a) The progress of globalization boosts demand for rapid cooperation, while technological advances that emerge in parallel and are obviously growing significantly improve digitization and digitalization.

b) There is a further targeted need to create instruments/tools that will increase the communication possibilities within globalization, facilitate cooperation, and accelerate and make this activity more effective, especially regarding information transfer. Hence, digital platforms and social media, among others, emerge first on a national scale and then on a global scale.

c) With the above-mentioned demand for the fast transfer, creation, and use of information, it has become possible to transmit information in an appropriate setting and with appropriate content or interpretation, which means that sometimes, in addition to reliable information, false or distorted information is also provided or created, opening a path for the emergence of today's fake news.

d) When fake news appears, there also emerges an immediate need for a defense system against it. This is to be provided based on appropriate interdisciplinary solutions that incorporate both technological and legal aspects.

As a result, digitization has become an effective tool for broadening the reach of globalization (standardization) and a primary transmission belt for information transfer and interpretation, which does not always correspond to the true intention behind the origin or dissemination of this information. The key, therefore, is on the one hand, to master the technology of creating, transferring, and disseminating information, and on the other hand, to create a defense system against false, modified, or misinterpreted information.

1.2 Digital platforms and social media operating in Poland

Globalization has also standardized the operation of digital platforms and social media. As previously argued, both have one concept in common: information. The need to and ease of spreading information today, to which globalization and digitization have contributed, have resulted in the emergence of specific tools that play a major role in creating, delivering, and interpreting information. These tools have taken the form of digital platforms and social media. These concepts are not

presented in the above order by chance. Digital platforms were created first, and improvements to the Internet, along with the development of advanced technologies, resulted in the emergence of the first social media.

From a technical point of view:

> A digital platform is a transmission medium of zero-one encoded television signals, data and voice, intended for direct reception and use by individual and collective recipients, and colloquially it is also a collection of content, such as television and radio programs, electronic publications, data, computer programs encoded in the zero-one system in the form of a bundle of compressed streams and organized by a single operator[9]

Therefore, the digital platform was initially identified with the effect of digitization in the domain of television.

In this context, digitization in Poland is mainly associated with:

> the launch of DVB-T digital terrestrial television, which took off in Poland in the second half of the 1990s. Its introduction was dictated by the more efficient use of frequencies and the offering of a new type of service, which analogue technology could not warrant.[10]

Historically, the first date in the digitization of terrestrial television in Poland is 1997, when the strategy for the launch of the DVB-T network was drafted. Subsequent important dates in this context in Poland include 4 May 2005, when the Council of Ministers signed a regulation titled the Strategy for the Transition from Analogue to Digital Terrestrial Television, and 11 December 2007, when Directive 2007/65/EC of the European Parliament and of the Council on audiovisual media services was enacted, linking these services with television broadcasting technologies and establishing a classification of services provided through analogue and digital television, Internet broadcasting and live streaming, and near video on demand.

The burden of deploying such changes to switching off analogue television and transforming it into digital television rested on a Polish constitutional body, the National Broadcasting Council. The process began in early June 2011.[11]

Currently, as we should only note for informational purposes, the digital platforms analyzed briefly in this paper include online platforms that appeared much later than digital television platforms. Therefore, "the concept of [the] 'online platform' can now be understood in various ways, for example, as identical to 'Internet website,' that is, a web page presenting a wide range of thematic content made

9 Bryndal and Kochański, 1998.
10 Myślak, 2019, p. 37.
11 Ibidem.

available to the user."[12] On the other hand, online platforms can also be defined as "a new business model of virtual intermediation between at least two distinct but interdependent (networked) user groups, being parties in multisided markets."[13] European documents indicate that online platforms share some important and specific characteristics. In particular, they have the ability to create and shape new markets, to challenge traditional ones, and to organize new forms of participation or conducting business based on collecting, processing, and editing large amounts of data; they operate in multisided markets but with varying degrees of control over direct interactions between groups of users; they benefit from 'network effects,' where, broadly speaking, the value of the service increases with the number of users; they often rely on information and communications technologies to reach their users, instantly and effortlessly; and they play a key role in digital value creation, notably by capturing significant value (including through data accumulation), facilitating new business ventures, and creating new strategic dependencies.[14]

Historically, the first Polish digital platform in the television domain was a platform named Wizja TV, which was launched on 18 September 1998 by the American company @Entertainment. On 16 November 1998, the digital platform Canal + started to operate.[15]

Currently (2021) in Poland, there are three popular digital platforms: Cyfrowy Polsat, Canal +, and Orange TV. Naturally, these operate mainly in the television domain; however, they offer a wide variety of media services.

The owner of the first one is the joint stock company Cyfrowy Polsat, which is one of the largest operators in Central and Eastern Europe. As a satellite TV operator, it is among the market leaders in terms of the number of subscribers across Europe.[16] In Q1 2020, the Cyfrowy Polsat Group posted revenue of PLN 2.85 billion (+ 2% per annum) and a net profit of PLN 184 million, while EBITDA amounted to PLN 1.027 billion (-1.1%). During this period, the Group's sales covered 466,000 new service contracts.[17] In 2020, the number of subscribers was approximately 5.55 million.[18] Cyfrowy Polsat's offerings include paid TV service, i.e., approximately 170 channels broadcast via satellite, terrestrial, and Internet technologies (IPTV, OTT); the provision of modern OTT services (e.g., Cyfrowy Polsat GO, PPV, VOD) and Multiroom, including online video services offered on the subscription and transaction models (PPV) (IPLA service); telecommunications services including voice and data transmission services; as well as various value-added services (VAS), broadband mobile

12 Wyrwińska and Wyrwiński, 2018, p. 97.
13 Śledziewska and Włoch, 2020, p. 99.
14 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 25 May 2016, COM (2016), p. 288.
15 Nowak, 2019, p. 251
16 See: https://bit.ly/3nYnwAh.
17 See: https://bit.ly/2VWlBjS.
18 See: https://bit.ly/3tZjb0H.

Internet services using modern LTE, LTE Advanced, and 5G technologies; television broadcasting and production services via Telewizja Polsat, offering 36 popular TV channels; and Internet media or wholesale services on the interconnection market, including, *inter alia*, interconnect services, IP and voice traffic transit, and line lease or domestic and international roaming services.[19]

Canal Plus is another platform on the market. It is a shared brand incorporating a satellite digital platform and Internet television services offered by the Polish company Canal + Polka S.A., a member of the French media group Groupe Canal +. The Polish Canal + platform was created as a brand replacing nc +.[20] Currently, Canal + Polska is a leading producer of premium thematic TV channels, offering a unique combination of premium segment programming, innovative technology, and a wide distribution network. The Canal + Polska Group serves over 2.7 million customers (as at 31 December 2020) as the operator of a satellite platform, offering TV packages that include both its own TV channels and third-party channels, with a particular focus on the premium segment.[21] It is the second largest distributor of paid TV packages in Poland, commanding a 21% share of the traditional paid TV market.[22] It is estimated that at the end of June 2020, Canal + Polska had 2,703 million customers. Their operating profit was approximately PLN 155 million, and net profit amounted to approximately PLN 121 million.[23]

The third platform is Orange TV, a digital platform providing telecommunications services in Poland, with a presence in all segments of the telecommunications market. The Group is the owner of the largest telecommunications infrastructure in Poland, providing voice and data transmission services on fixed and mobile networks. As one of the leading telecommunications operators in Europe, Orange S.A. owns 50.67% of Orange Polska's shares.[24] According to publicly available data, Orange Polska's entire base of combined package users totaled 1.387 million. The company reaches 4.4 million households in 147 cities. In 80 cities with optical fiber infrastructure, it reaches more than half of the households. Orange Polska's revenue for Q1 2020 increased by 0.9% to PLN 2.804 billion, and the operating profitability ratio EBITDA increased by 6%, up to PLN 676 million, compared to the same period in 2019.[25]

As the other element of the fake news phenomenon, our analysis will cover social media or social networks, which play the same role here as digital platforms, as they are the relevant domains for the creation, transfer, and dissemination of fake news. In Polish, the term is "a direct translation of the English term 'social media' or 'social

19 See: https://bit.ly/3AtIc6x.
20 See: https://bit.ly/2Zhmp4i.
21 See: https://bit.ly/3EEV9Nm.
22 See: https://bit.ly/3kq3WKL.
23 See: https://bit.ly/3AsZhO1.
24 See: https://bit.ly/3lIJu7l and https://bit.ly/3nSEuQv.
25 See: https://bit.ly/3zu9uIH.

networks.'" [26] The term 'social networks' first appeared in the United States in the 1950s. Historically, the beginnings of social networks, strictly in today's sense of the term, date back to the 1990s.

In an attempt to explain what a social networking service is, it should be noted that the author of the concept is commonly agreed to have been Professor John Barns, who defined a social network as a group of approximately 100–150 people who share an interest in the same task, job, or hobby.[27] In 1995, in the United States, Randy Conrads started a service under the name Classmate.com. The creator of the service set the goal of building a network among people who had once kept in contact with each other so that they could exchange information, strengthen or renew relationships, etc. Similar initiatives were mirrored in Poland with the appearance of the service Naszaklasa.pl, which was created by computer science students at the University of Wrocław, namely Maciej Popowicz, Paweł Olchawa, Michał Bartoszkiewicz, and Łukasz Adziński. It is worth noting that the social networking site Epinions.com was launched in 1999, and in 2003, Tom Anderson and Chris DeWolfe created MySpace.com.

There is no uniform and universally binding definition of social media in the legal literature or in Poland's legal system.

However, it is indicated that "the elements shared across the definitions of social media are: creation of information and multimedia content, texts, photos for personal use and dissemination of the same among friends."[28] Social media are often defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content."[29] It is accepted that:

> Media as such are carriers of information (media, tools for recording and transmitting information); they can perform two elementary functions: i.e. information functions – they publish news and all kinds of references to these (opinions, comments, debates, polemics, etc.). Therefore, they are content carriers that serve directly to provide information, acquire and expand knowledge (presenting and describing the world), and entertainment functions — they provide the opportunities of spending time pleasantly.[30]

On the other hand, the characteristic features of social media are as follows: they can be used on any scale; the means of production are available to everyone interested; the publication of information is only the beginning of the media process; the original information can be modified infinitely; access to the creation and reception

26 Sudomir, 2020, p. 97.
27 See: https://www.historyofinformation.com/detail.php?id=715.
28 Tomczak, 2017, p. 145.
29 Kaplan and Haenlein, 2010, pp. 59–68.
30 Kaznowski, 2010.

of content is free; without social participation, the idea of social media cannot be pursued; the final value of information is directly influenced by the participation of the social group (community) that is actually focused around the topics discussed; each interested party has access to their own and other contributors' content at any time and has the option to refer not only to the underlying content but also to the contribution of other authors; no coordination between authors; no elements resulting from the creation (co-creation) process are deleted and they are continuously available; the content is spread through social interaction (which directly translates into the scale of distribution of each piece of information); the delay between the creation of content and its publication is kept to a minimum (no delay); unforced way of content creation.[31]

It is an uncontroversial fact today that:

> Social media are an important part of our everyday life. They emerged as a contemporary response of the digital world to the primordial human need, which is the need for social contact, as well as the need to connect into social groups. Social networks are one of the most popular communication tools on the Internet.[32]

There is also common agreement that social media have a communication and information function, mainly serving to facilitate the exchange of experiences, opinions, and views.[33]

In technical terms, social networking services are classified as Web 2.0 generation media, i.e., a group of media existing and operating online, where the users are in fact responsible for the content posted, and thus are both users and creators.[34] Hence, it is emphasized that:

> Web 2.0 is an approach to communication on the Internet, which takes into account a change in the position of the recipient, who also becomes a full participant in the dialogue. In Web 2.0, the consumer of content also becomes its producer. Web 2.0 is therefore based on participation via the Internet (…)

This term covers "Internet services that allow users to collaborate and exchange information online through social networking sites."[35] The emergence of Web 2.0 was a major breakthrough, as the previous generation, i.e., Web 1.0, was only a

> one-way communication model, in which the content posted on various websites was primarily managed by adequately qualified message creators. The recipients

31 Ibidem.
32 A. Bąk, 2016, p. 139.
33 Delińska, 2018, p. 19.
34 Gogołek, 2010, p. 160.
35 Flasiński, 2017, p. 175.

> themselves were passive. They could only read the information posted without any options to build or comment on it.[36]

As is clear, Web 2.0 demonstrates users' multisided joint cooperation and complementarity.

Returning to terminological considerations, however, there is currently no agreement as to the definition of a social networking service. The easiest way is to begin to analyze the structure of the term 'social media.' In Polish usage, it is a direct translation of the English term 'social media.' In this translation, 'social' signifies the social element, and 'media' means an information carrier.[37] However, there are no consistent, unambiguous, and relatively simple definitions of the term, nor is it defined by law in the Polish legal system, hence the different terms denoting social media in Poland, such as 'social networks' or 'social networking services,' etc. Social media are defined as, *inter alia*:

> an information service on a computer network, publicly and commonly available at a single WWW address, presenting content of interest to all network users, featuring optional, specialized online functionalities (e.g. news, chat, online discussion forum, free e-mail, web hosting, internal and external search options via a search engine).[38]

In dictionary terms, social media or networks are defined as an online service co-created by a community of Internet users with similar interests that allows them to contact friends and share information, interests, etc.[39] It is accepted that "the main operating principle of social networks is to enable building users' own, private or public personality profile, where specific information about a person, company or organization is posted."[40] Elsewhere:

> The term social media most often denotes a set of tools based on online media and mobile technologies that enable the exchange of information in the form of an interactive dialogue between users, bypassing the limitations related to, *inter alia*, the place of residence.[41]

A very broad interpretation of this term states that even "every page on the Internet on which users interact is a social medium."[42]

Whenever a social medium exists, its natural environment is the Internet. Access to the Internet is the basis for social media's reach and degree of interest.

36 Sarowski, 2017, p. 34.
37 Dziwulski and Ogrzebacz, 2017, p. 87.
38 Tytko, n.d.
39 See: https://sjp.pwn.pl/sjp/serwis-spolecznosciowy;5579205.html.
40 Donecki, n.d. Available at: http://www.publikacje.edu.pl/pdf/11046.pdf.
41 Wicińska, 2017, p. 115.
42 Czaplicka, 2014, p. 10.

According to publicly available data as at January 2021, 31.97 million people use the Internet in Poland. This accounts for approximately 84.5% of Poland's total population. The same source indicates that the average Pole (aged 16–64) spends 6 hours 44 minutes on the Internet per day, including 2 hours on social media. For comparison, the average duration of television consumption in Poland is about 3 hours 15 minutes, and that of online and printed press (combined) is 1 hour 16 minutes.[43]

The above data show the Internet's enormous power at present. Its status surely warrants a review and characterization of the individual social media sites operating in Poland. It seems that there is no need to describe the exact profile of the selected popular social media in Poland, as, generally speaking, these are globally recognized entities. The only social medium that specifically operates in Poland is Albicla; it is an entity that has just started its activity and precise data for it are not currently available.

However, in the analyzed context, the data related to the operation of these services/sites in Poland are interesting, so let us examine, as far as possible, the domestic landscape.

The YouTube service is perhaps the best starting point for a review of social media in Poland. At present, it is the most popular social media site in Poland, used, on average, by approximately 92.8% of users. The site reaches over 24 million people in Poland. Statistics show that Polish women spend an average of 40 minutes on YouTube.[44] YouTube's viewership in Poland is over 91% of Polish Internet users. As the data demonstrate, out of this number, 10 million YouTube viewers are aged 24–44, and in total, YouTube reaches 24.6 million viewers in the country.[45]

In terms of popularity, Facebook is next. Its community includes approximately 89.2% of Polish Internet users. More in-depth data indicate that approximately 96.6% of these users use Facebook on their mobile device. The average user posts at least one like per day, and three comments in a 30-day period (with women leaving comments as much as five times more often than men, who comment twice, on average, over the same period). The overall community of Polish Facebook users numbers approximately 18.3 million people, with the largest group among them belonging to the 25–34 age group (27.8%).[46] Facebook Messenger, which operates as an independent platform, has been installed by approximately 76.5% of the community, which is about 16,018,455 users.

Third in the popularity ranking is Instagram, with 60.6% of Internet users, meaning that in Poland, it is used by nearly 9.2 million people. It is followed by Twitter, with a 37.5% user base or 1.35 million people in Poland.

43 See: https://empemedia.pl/social-media-w-polsce-2021-nowy-raport/.
44 See: https://bit.ly/2XxaHl5.
45 See: https://spidersweb.pl/2020/11/youtube-polska-statystyki-2020.html.
46 See: https://www.whysosocial.pl/uzytkownicy-social-media-w-polsce-i-na-swiecie/.

Next is LinkedIn, with 24.6%, corresponding to 4.10 million users in Poland. The service is popular among middle and senior management, as according to statistics, on average, 97% of managers using social media have reported using LinkedIn.

It is noteworthy that the most popular sites in Poland include Snapchat, with 28.9%, and TikTok, with a 28.6% user base.[47]

Social media as an information source:[48]

Country	Score
Greece	74%
Brazil	72%
Hungary	64%
Poland	58%
Denmark	56%
Turkey	73%
Portugal	66%
Spain	60%
Sweden	56%

2. An attempt to determine the scale of influence, benefits, and dangers of digital platforms' and social media's existing operating structure

At the outset, one should agree with the statement that:

> For people today, technological progress in the field of social media saves time and money, and also facilitates everyday activities, communication and interpersonal contacts. Like any invention, innovation or advanced solution, however, social media also carry risks, dangers and negative effects.[49]

This statement very accurately reflects the present situation, in which social media and digital platforms are key features. While digital platforms are intended

47 See: https://empemedia.pl/social-media-w-polsce-2021-nowy-raport/.
48 See: https://biznes.newseria.pl/files/raport-fake-news-newseria.pdf.
49 Stecuła, 2017, p. 230.

for the presentation of information, entertainment, educational content, etc., usually associated with a specific decision making place (editorial office) where the content is prepared in advance, social media operate on their own rights, as—being entities based on the Web 2.0 philosophy—they allow active content management by those who create it from scratch, remake existing content, or become transmitters of information created by other actors. Content transfer is very dangerous; while digital platforms have a permanent entity owner, an editorial office, or certain action plans, there are no such fixed elements with social media. For social media, the user only utilizes the tools an entity creates and offers to independently generate or promote specific content.

It is impossible to catalogue the greatest social media-related risks, as this would depend on the area and direction selected for the purposes of the analysis. Therefore, it is not possible to present a specific catalogue of these risks, or even benefits. This depends precisely on the area of operation and the target profile of a specific entity falling within the scope of the term 'social media' or 'digital platform.'

Therefore, if one is to analyze the fake news phenomenon, the obvious area of interest is mainly information, and solely in this context, it is worth considering the general risks associated with the inextricable links between the terms 'information' and 'social media.' There is also no doubt that it is much easier to present the general advantages of social media than their disadvantages. The advantages include, *inter alia*, ease of communication, rapid access to information, fast information sharing, the opportunity to learn about various types of information sources, etc.[50]

As mentioned above, information is at the core of the fake news phenomenon. The data quoted above representing the number of digital platform and social media users demonstrate the enormity of the scale of influence. These data demonstrate that in Poland, like in other countries in the world, the number of users is counted in millions. This translates into the huge influence these entities have on community members. There is no doubt that these entities can use their power of influence in various ways and not necessarily for the common good. Therefore, in every country, a security system is extremely important to ensure protection against the promotion of vast amounts of content and information through these entities, which everyone will naturally consider negative. We assume that "information transferred or used is or should be based on reliability, understood as [a] full-fledged, credible source of information and truth as the essential content of information, being consistency of thought with its object."[51]

In view of the above, it seems that the primary negative influence digital platforms and social media exert is the planned or incidental creation and dissemination of untruths, or simply put, falsehoods, both these concepts denoting 'lies'

50 Jankowski, 2019, p. 268.
51 Dębowski, 2014, pp. 12–15.

(mendacium) and thus *locutio contra mentem*, i.e., 'speaking contrary to one's mind,' that is, a statement inconsistent with one's conviction.[52]

If one of the activities of social media is related to information, the opposition to this term is disinformation distributed by entities using their reach and technological capabilities.

Information and its creation, transmission, and interpretation generate interest among the general public. Further, this interest generates authority. It turns out that it is those who transfer their own content or content prepared by others via a digital platform or social media that often become an authority, i.e., a person or an institution enjoying particular recognition.[53] The distribution of disinformation entails a kind of overturn in the hierarchy of authorities as regards knowledge, interpersonal relations, state authorities, etc.[54] This conversion of authority consists of creating, transmitting, and commenting on information, news, or data in a way that is contrary to the truth. In this way, it is possible to subvert the natural axiology of things and, through false actions, lose natural values in favor of anti-values. Confronted by a flood of information, a person must evaluate and segregate it, without knowing which pieces are true and which are false. Information, in turn, usually reaches us after it has been captured and processed by algorithms created in a predetermined model and directed to perform a specific action. Mastering algorithms seems to be key.

Further, it is not without significance that to check the credibility of information is to investigate its sources, and this requires a considerable amount of activity and intellectual effort, which are quickly declining in today's society. After all, information is what has been said or written about someone or something, and the communication of something.[55] Information creates reality, gives an edge, and resolves many issues. Reality creation motivated by untrue information cannot be allowed to trigger other even more harmful activities. It is also important that the proposal address "ethical and axiological dilemmas relating to communication via, for example, social media, and . . . talk more about the need for ethics in these media (...)."[56]

Perhaps the catalyst shielding us from dangers lies simply in ethics and axiology. For this, it is necessary to understand these terms as they are—a task which is very difficult at present. It is difficult because the multifaceted evolution of human civilization, accompanied by scientific and technical progress, the multiplication of human expectations, and the persistent emergence of needs and options to meet them, has put a very strong hold on the ethical and moral attitudes that have proven effective for centuries. This is especially visible with information, as it is not uncommon that information based on truth cannot penetrate the public domain, while false information is immediately propagated as simply more attractive.

52 Wolniewicz, 2012, p. 5.
53 See: https://sjp.pwn.pl/sjp/autorytet;2551342.html.
54 Werner and Trzoss, 2019, p. 148
55 See: https://sjp.pwn.pl/sjp/informacja;2466189.html.
56 Laskowska, 2012, p. 9.

Humans should act ethically and morally; that is, we should try to apply standards of the highest order to ourselves and act in line with our conscience toward others. A moral person is an individual that adheres to specific principles that have been set and which operate in human communities to allow for the distinction between good and evil and between proper and improper conduct. Morality contains the characteristics of truthfulness, credibility, and humility.

Therefore, it seems that the only panacea to challenge the falsehood that underlies fake news and thus poses grave danger is returning to the basics of human existence by returning to natural law. Disinformation is wrong at its roots, as in the context analyzed here, it promotes untruth and falsehood or anti-values. The response lies in the principles of natural law, since

> from the philosophical point of view, natural law allows us to establish that law exists in human nature, to know its nature and significance, and thus to realize that it is a criterion that enables us to distinguish good from evil, determines the principles of conduct and the strength of the moral obligation under positive legal norms[57]

A return to the natural system of values—understood as something absolute that sets the direction of positive action[58]—seems to be the key to controlling the current negative influence of the fake news phenomenon. Values should be the basis for designing new legislation concerning the operating domain of fake news.

3. A review of national legislation for the admission of digital platforms and social media to individual country markets (organizational form, country branch office, legal obligations, operating restrictions, etc.)

Let us now focus our analysis on a review of the legislation under which digital platforms and social media operate in Poland.

At the level of European legislation, there is currently a debate over the Regulation of the European Parliament and of the Council on the single market for digital services (i.e., the Digital Services Act, DSA). Briefly, the act sets out to improve the functioning of the digital single market and ensure effective supervision over service providers operating on the Internet; enhance security and protect freedom of expression online; increase the transparency of the operation of online platforms, e.g., for Internet advertising or content moderation; ensure that very large online platforms act responsibly in order to limit the risks arising from the use of their

57 Laskowski, 1991, p. 151.
58 Wielec, 2017, p. 32.

services ('very large online platforms' are those whose services are used by at least 10%, or 45 million users in the European Union [EU]).[59]

Poland's national legislation classifies digital platforms as entities of economic law that are subject in the first place to laws and regulations as any economic entity.

Hence, each of the digital platforms described above is a commercial company. These are joint-stock companies incorporated and existing under the Code of Commercial Companies, where the joint-stock company in the Polish legal system is defined as:

> a body corporate whose structure consists of members who, through the contribution of shares, set up the assets of the body corporate and under their rights and obligations direct its activities. A joint-stock company is a capital society (organization) with a varying personal composition and having its own assets.[60]

In addition to the legislation on companies, there are also a number of additional legal acts that regulate business activity in Poland and provide the basis for the operation of digital platforms.

An interesting issue here is the supervision of these platforms' activities. Firstly, it should be noted that under the system of Polish law, supervision takes the form of constitutional control. The Constitution of the Republic of Poland of 1997 provides for a body referred to as the National Broadcasting Council. Specifically, according to Article 213 of the Constitution, the National Council of Radio Broadcasting and Television safeguards the freedom of speech, the right to information, and the public interest regarding radio broadcasting and television. The National Council of Radio Broadcasting and Television issues regulations and, in individual cases, adopts resolutions. In organizational terms, the Council's members are appointed by the Sejm, the Senate, and the president of the Republic. A member of the National Council of Radio Broadcasting and Television may not belong to a political party or a trade union, or perform public activities incompatible with the dignity of their function. On the other hand, the rules and procedures of the National Broadcasting Council, its organization, and detailed rules for appointment of its members are specified in a statute, i.e., the Broadcasting Act of 29 December 1992. According to these provisions, the National Council safeguards freedom of expression in radio and television, the independence of media service providers, and the interests of recipients, and ensures the open and pluralistic nature of radio and television broadcasting. In this context, the Council's tasks include: 1) to draw up, in agreement with the Prime Minister, the directions of the State policy in respect of radio and television broadcasting; to determine, within the limits of powers granted to it under this Act, the terms of conducting activities by broadcasters; 2) to make, within the scope set forth by the Act, decisions concerning broadcasting licences to transmit and retransmit programme

59 See: https://bit.ly/3zxF2gx.

60 Sołtysinski, 2016. Available at: https://bit.ly/3AtJ9f7.

services, entry in the register of programmes, hereinafter referred to as the 'register', and to keep the register; 3) to grant to a broadcaster the status of a social broadcaster or to revoke such status, on terms laid down in the Act; 4) to supervise the activity of broadcasters within the limits of powers granted to it under the Act; 5) to organise research into the content and audience of radio and television programme services; 6) to monitoring the market of on-demand audiovisual services in order to identify the group of entities providing on-demand audiovisual services and assess the performance of their obligations under the Act; 7) to determine fees for the award of broadcasting licences and registration; 8) to determine licence fees in accordance with the principles set forth in the Licence Fees Act; 9) to act as a consultative body in drafting legislation and international agreements related to radio and television broadcasting or on-demand audiovisual services; 10) to initiate research and technical development and training in the field of radio and television broadcasting; 11) to organise and initiate international co-operation in the field of radio and television broadcasting, including co-operation with regulatory bodies of Member States of the European Union competent for radio and television programme services; 12) to co-operate with appropriate organizations and institutions in respect of protecting copyright as well as the rights of performers, producers and broadcasters of radio and television programme services; 13) to initiate and supporting self-regulation and co-regulation concerning the provision of radio and television programme services; 14) to promote media literacy (media education) and to co-operate with other state bodies, non-governmental organizations and other institutions in respect of media education. The National Council consists of five members, of which two are appointed by the Sejm, one by the Senate, and two by the president, from among persons with a distinguished record of knowledge and experience in public media. The chairman of the National Council is elected and dismissed by the Council from among its members. Upon a motion from its chairman, the National Council elects a vice-chairman from among its members. Council members' term of office is six years from the most recent member's day of appointment. Council members perform their functions until the appointment of their successors. A member may not be appointed for another full term of office. The body empowered to appoint members dismisses members solely in cases when the said person has resigned; has become permanently unable to discharge of duties for reasons of ill health; has been sentenced for a deliberate criminal offence and the said sentence is valid and enforceable; or has submitted an untruthful screening statement, as confirmed by a final and valid decision of the court; or has committed a breach of the provisions of the Act and the said breach has been confirmed by the decision of the Tribunal of State.

The situation is completely different when it comes to social media. In the Polish legal system, there is, so far, no law dedicated to the organization and operation of social media. Therefore, these are mainly governed by EU legislation and general legal principles often derived from constitutional rules.

One of the few acts with a certain degree of influence on the social media market is the Electronic Services Act of 18 July 2002, which specifies *first*; obligations of the

service provider related to the provision of electronic services; *second*; rules for excluding the service provider's liability for the provision of electronic services; *third*; rules for the protection of personal data of natural persons using services provided electronically. The Act lays down definitions of a number of terms, including *first*; providing services by electronic means, being such a way of rendering a service, which comprises transmitting and collecting data by means of electronic processing devices, including digital compression and data storage systems, at the individual request of a service recipient, without the parties being simultaneously present (remotely), while the data are transmitted through telecommunications networks; *second*; electronic communication means, being technical measures, including teleinformation equipment and software tools co-operating with it, enabling individual distant communication by using data transmission between teleinformation systems, in particular electronic mail; *third*; service provider, being any natural person, body corporate or organizational unit without legal personality, who, while performing, even as side activities, commercial or professional activities provides services by electronic means; *fourth*; service recipient, being any natural person, body corporate or organizational unit without legal personality, who uses services provided by electronic means. The Act also contains penal provisions, under which any person who fails to submit or submits false or incomplete data is liable to fines, and any person who transmits unsolicited commercial information by electronic communications means is liable to fines.

The above legal acts constitute the general core of the legislation concerning the operation of digital platforms and social media.

As indicated at the outset, there is currently no single act in the Polish legal system that comprehensively organizes the functioning of social media specifically.

4. The concept of fake news

The phrase 'fake news' was borrowed into the Polish language from English. The term is made up of two words, of which the first means, in the Polish translation and understanding, falsehood, imitation, counterfeit, forgery, fraud, deception, or fabrication, while the second, means recent or new events, information, intelligence, or report. The combination of these two terms is quite specific, as while 'fake' is by definition a negative concept associated with something wrong (false, fraud, etc.), 'news,' meaning information, is neutral.

4.1 Dictionary terms

The term 'fake news,' apart from the above etymology, does not have a binding definition in Polish law. To be precise, it does not have a legal definition that is often

employed in various jurisdictions. Polish law makes frequent use of legal definitions, which are deemed to be "a statement by the legislator that specifies the sense and meaning of a word or expression being defined, or gives an unambiguous characteristic of the object being defined."[61] It is rightly pointed out that:

> The legal definition is one of the legislative measures used in the law-making process, aimed at clarifying a concept used in the text of a normative act, and thus at facilitating the understanding of a legal norm in accordance with the intention of the legislator.[62]

4.1.1. 'Word of the year' designation.

Nevertheless, none of the above acts concerning the operation and organization of digital platforms or social media contains a legal definition of 'fake news' or any other juridical definition of the concept. A legal act with a legal definition of 'fake news' would certainly be a very positive step forward. For the time being, however, the only available option is to define this concept on a doctrinal basis.

Inter alia, it is emphasized that the term 'fake news' is a neologism with no formal definition. In rough translation, one can say that this is a message intended to mislead the recipient. It is neither truth nor a lie. Fake news is usually based on disinformation or a prank, often containing elements of truth. "Fake news can pretend to be real information, articles, social media posts, memes, etc. It can be created with a variety of intentions, ranging from fraud, propaganda tools, [or] sensationalism, to a prank."[63]

The term was singled out as the Collins Dictionary Word of the Year 2017 due to its 'ubiquitous use,' marked by a 365% increase in usage frequency over the several months prior to its 'word of the year' designation. According to the editors of the Collins Dictionary, the word combination 'fake news' is 'ubiquitous' and extremely popular.[64]

In Poland, the term 'fake news' was submitted as a candidate for the Youth Word of the Year 2017, which is a ranking organized by Wydawnictwo Naukowe PWN scientific publishers in cooperation with the Key Words project as part of the National Centre for Culture Poland initiative Native Tongue – Add to Favourites.[65]

In dictionary terms, 'fake news' is defined as untrue or false information most often disseminated by tabloids with a view to causing controversy or slandering or libeling someone (usually a politician).[66]

61 Malinowski, 2005, pp. 215–216.
62 Bąkowski, 2017, p. 57.
63 See: https://cik.uke.gov.pl/news/fake-news-czyli-falszywa-prawda,191.html.
64 See: https://tvn24.pl/kultura-i-styl/slowo-roku-2017-fake-news-ra787106-2483140.
65 See: https://sjp.pwn.pl/mlodziezowe-slowo-roku/;202298;3.html.
66 See: https://sjp.pwn.pl/mlodziezowe-slowo-roku/haslo/fake-news;6368870.html.

To some extent, it is a neologism that is extremely difficult to frame in definitions, as it can mean a media message that is neither true nor untrue and is based on disinformation, though often containing elements of truth.[67]

'Fake news' also denotes information that may have multiple significant financial or political implications.[68] There is common agreement that:

> Technological progress, broadly understood globalization, the growth of the Internet and social networks, the relativization of the truth, cultural and moral changes, the race to be the first to publish information, and the resulting decline of reliable journalism, overstimulation of consumers, a decline in trust in media institutions — these are just a few factors that have made fake news triumph at this point.[69]

'Fake news' also describes "individual posts, messages or even entire news channels where the transmitted data (to varying degrees) turn out to be false or distorted."[70] Therefore, "the concept of fake news is often referred to as various cases of information manipulated or tampered with by authors/broadcasters."[71] Further, fake news is said to be "false, often sensational news, disseminated as an objective information message."[72] Elsewhere, it is emphasized that "the concept of fake news is most often defined as misrepresentation, often of a sensational nature, published in the media with the intention to mislead the recipient for financial, political or prestigious benefit."[73] In another approach, fake news refers to post-truth. This view highlights that:

> Another form of post-truth that thrives in social media is the fake news. Its power is driven by the emotions of the recipients; hence it is often based on religious beliefs, values, stereotypes or bias. For fake news to be effective, it must refer to some concept that already exists.[74]

According to a complementary approach:

> 'Post-true' content is called fake news. Its popularity among recipients is driven not by facts but emotions, therefore it is often based on religious beliefs, values, stereotypes, prejudices, etc. In order for fake news to be effective as a tool of mass persuasion, it must refer to concepts already existing in the consciousness of some social group. Otherwise, it would take a long-term process and mutual effort to build ideas

67 Bąkowicz, 2019, p. 281.
68 Woźniak and Zapór, 2017, p. 100.
69 Podlecki, 2017, pp. 128–129.
70 Waszak, 2017, p. 175.
71 Daniel, 2018, p. 99.
72 Brenda and Mańkowska, 2019, p. 11.
73 Wójcik, 2019. Available at: https://bit.ly/39mgVXR.
74 Flader, 2018, p. 52

from scratch in the minds of recipients."[75] It is noted that "fake news, as information made up by Internet users, travels *en masse* and instantly on the web, especially with the help of social networks.[76]

For the purposes of a broad description of the concept, one can use any of the approaches, according to which:

Fake news is: (1) a false message having the characteristics of a true one; (2) a satirical message created deliberately for entertainment purposes; (3) which one thinks is true, but [which] is [actually] false; (4) designed to mislead for financial, political and prestige gains; (5) a false message, regardless of the intention of the sender.[77]

5. Classifying fake news

Classification is an ordering operation guided by a predetermined criterion.

Among the current fake news classifications, fake news is categorized as *first*; satire or parody, *second*; false combination of a headline, image or caption incompatible with the content, *third;* misleading content, *fourth;* false context, *fifth;* fraudulent content, *sixth;* manipulated content, *seventh;* fabricated content.[78]

Moreover, fake news can be divided into 1) intentionally untrue fake news, satirical 'with a pinch of salt' (satire news, etc.), sometimes referred to as truthiness; 2) fake news which imitates real news but is completely fabricated; 3) fake news created on the basis of real news, subjected to manipulation; 4) real news referred to as fake only because someone did not like it.[79]

All the above classifications, however, give criteria that are not related to the legal system, which is problematic because our analysis sets out to define them in legal terms. Though these divisions may be interesting, they are from the point of view of social science or society itself and do not necessarily reflect the legal effects.

Therefore, when classifying fake news, one can—using the criterion of its power of influence in conjunction with its scale of liability under the law—propose the following categorization/divisions in the legal domain:

75 Bakalarski, 2017, p. 12.
76 Łyszczarz, Sierocki and Sokołowski, 2018, pp. 5–6.
77 Palczewski, 2019, p. 17.
78 Bąkowicz, 2019, p. 285.
79 Palczewski, 2017, p. 31.

First degree fake news would represent the gravest misrepresentation, false information, and content load, with the biggest impact. Examination would cover who and what such fake news concerns and whether the creation and publication of fake news threatens the highest values enjoying protection under the law, such as public order, the health and life of citizens, etc. In this case, such action should be penalized under the criminal law. In other words, the construction of an appropriate criminal provision should be envisaged, which would penalize fake news as a cause of action with enormous effects on multiple levels. Examples would include alteration, manipulation, falsification of indisputable historical facts or presentation of the course of a certain event carrying a huge social load in a manipulated, falsified manner with a view to misrepresenting, ridiculing or discrediting key historical facts or state leadership, or creation and dissemination of such information that will endanger public security. However, the constituent elements of the crime must be precisely defined and should include, *inter alia*, the intent and purpose of creating or disseminating fake news.

Second degree fake news would be an act with a much more limited impact, affecting more the repute of a person or a fact and violating only the private area of the person or fact. There is no major impact on the public, but the entity that is the main subject of fake news is discomforted. In this case, the message is so satirical or distorted that, in principle, any reasonable bystander would point to a significant transgression of, for example, aesthetic or moral norms. Although private interest is violated, it is not necessary to use penal measures. Therefore, the best domain of legal liability is civil law or civil action, along with the use of any tools that exist even now (court action, redress, etc.). In this case, it is under civil law that all issues related to the creation, dissemination, and use of fake news will be resolved.

Third degree fake news would be the use of manipulated content or false information within one's professional group. This is a much more limited area of impact than that indicated above. An important factor here is the professional or social group, which functions according to generally established principles of professional deontology. In this case, disciplinary/professional liability comes into play, i.e., a type of liability reserved for a specific group of entities.

Fourth degree fake news would finally be a minor, essentially non-punitive, formal-only (i.e., non-consequential) production and dissemination of fake news for satire, fun, etc., without a major consequence for such production, distribution, or use. This type of fake news is not penalized in any way.

Of course, the classifications presented above are only proposals, as to have this idea signed into law would require strong legislative support, targeting amendments to various areas of law. However, these are proposals that refer directly to the criterion of legal liability.

6. Fake news and deep fakes – differences and similarities

Fake news research cannot ignore a specific type of fake news called the 'deepfake,' which has, in terms of social correlations, emerged as a special type. There is undoubtedly a close correlation between fake news and deepfakes. It is rightly emphasized that "deepfakes are a breakthrough innovation that sets new frontiers of human cognitive abilities in a digital environment, a technology that is used for various purposes, from '(video) hate speech' to laudable social campaigns (...)."[80] It is also indicated that:

> It is an image synthesis technique based on artificial intelligence. It is used to combine and overlay existing images and videos onto source images using a special machine learning technique. Deepfake is a human imaging technology that uses artificial intelligence to alter human images.[81]

Deepfakes are simply information communicated using artificial intelligence (AI) that can combine, replace, and overlay images and video clips to create fake videos that appear authentic. Deepfake technology can depict someone's behavior in a video in a humorous, serious, emotional way, etc., but the acting takes place without the person's consent.

> Deepfake is always about a certain load of information presented for a predetermined purpose, using artificial intelligence, which, through its capabilities, enables distortion, creation, modification, etc., of the entire human figure, including its image, manner of movement, gestures, or distortion of any other type of information carrier, etc. As a result, highly credible information is created, e.g. a very realistic but also highly manipulated and false video, presenting activities that did not and do not take place in real time. One must agree that the use of 'deepfake' implies the use of machine learning algorithms and face mapping technology to digitally manipulate people's voices, bodies and faces.[82]

A deepfake prepared by professional experts will be a very sophisticated artefact, e.g., a distortion of voice and image used to convey information falsely. If the combination is perfect, the result will be a false reality devoid of any flaws that is essentially indistinguishable from the original.

80 Kreft, 2020.
81 Dąbrowska, 2020, p. 90.
82 Wasiuta and Wasiuta, 2019, p. 20.

7. Reasons for and ways of creating fake news and the role and importance of creating, disseminating, and using fake news

Due to the extremely rich diversity of the impact fake news exerts, it is impossible to present a complete and accurate catalogue of the reasons and methods for its creation.

Nevertheless, studies conducted in Poland suggest that the incidence of fake news will continue to rise, as indicated by as many as 91% of respondents. However, regarding the main reasons for the creation and dissemination of fake news, the same study reports that these include: a) progressing tabloidization of the media, which consists of the transformation of news into strongly simplified formats, capturing the recipient's attention through the use of flashy and controversial headlines, captions, and graphics;[83] b) competition in the information search market; and c) shortage of time to verify the information received due to the specific nature of today's times.[84] The response scores for the first, second, and third reasons were 67%, 66%, and 53%, respectively.

To generalize, it can be assumed on the basis of the doctrine that fake news is created and disseminated essentially for all sorts of reasons and in different domains, including political, financial, ideological, and advertising, as well as for entertainment, fun, attention, or publicity. It is often emphasized that fake news serves a) to increase the click-through rates, followed by specific financial gains; b) to manipulate on various levels, whereby it takes the form of a negative action, which undoubtedly poses very dangerous consequences; c) to commit strictly criminal acts such as fraud or phishing through linking to fake news; d) to promote trolling,[85] i.e., an act that is evidently "anti-social conduct characteristic of online groups, discussion forums, chat rooms and social networks, consisting in intentionally influencing other users in order to ridicule or offend them by sending offensive, controversial, often untrue messages."[86]

83 Całek, 2013, p. 312.

84 The survey was carried out using the CATI method on 24-28 April 2017 by the ARC Rynek i Opinia research institute. One hundred and fifty-four journalists participated in the study. Available at: https://bit.ly/2XJ1s1H.

85 See The Information Security Doctrine of the Republic of Poland. Available at: https://bit.ly/3nPx-FiB.

86 See: https://www.sunrisesystem.pl/blog/2180-po-co-komu-fake-newsy.html.

8. Legal liability for the creation, dissemination, and use of fake news (criminal liability)

Criminal liability carries a specific load of arbitrariness and authority on the part of public administration bodies in relation to the individual's legal position. Certain behaviors should always only be penalized under the criminal law regime as *ultima ratio*. Contemporary growth of the fake news phenomenon has not been met with any response, whereby the Polish legislator would decide, at this stage, to construe criminal law provisions dedicated to fake news. In the criminal law system in Poland, there is no provision that would provide for direct criminal liability for the creation, dissemination, use, etc., of information considered to be fake news. However, this does not mean that it is impossible to find indirect Polish criminal law provisions that would trigger criminal liability for broadly understood false information.

Firstly, it should be noted that the Criminal Code is the basic legal act in the Polish criminal law system.[87]

Where fake news takes the form of stalking, discrediting, or harming the good name of a person, etc., the above legal act provides for the offense of persistent harassment. Pursuant to Article 190a of the Polish Criminal Code, any person who by persistent harassment of another person or their next of kin evokes in them a justified sense of threat, humiliation, or distress, or significantly violates their privacy is subject to the penalty of deprivation of liberty for a term of between six months and eight years. Any person who pretends to be another person and uses their image, personal data, or other data serving their public identification with a view to causing them property or personal damage is liable to the same penalty. If an act specified above results in a suicide attempt of the person, the offender is liable to the penalty of deprivation of liberty for a term of between two and twelve years. The offense is prosecuted upon the aggrieved party's complaint.

If the fake news involves presenting, for example, a person, institution, etc., damagingly, then it is an offense of defamation. In this case, according to Article 212 of the Criminal Code, any person who slanders another person, a group of persons, body corporate, or an organizational unit without legal personality for conduct, or characteristics that may discredit them in public opinion or result in a loss of confidence necessary to perform duties in their position, occupation, or type of activity is liable to fines or the penalty of restriction of liberty. If the offender commits the above act through the mass media, they are liable to fines or the penalty of restriction or deprivation of liberty for a term up to one year. The offense of defamation is also subject to private prosecution.

A similar regulation can be found in Article 216 of the Criminal Code concerning the offense of insult, under which any person who insults another person in their presence or publicly in their absence, or with the intention that the insult will reach

87 Criminal Code Act of 6 June 1997, Journal of Laws No. 2020, item 1444.

the person is liable to fines or the penalty of restriction of liberty. It is also provided that any person who insults another person using the mass media is liable to fines or the penalty of restriction or deprivation of liberty for a term up to one year. Private prosecution also applies in the case of this offense.

Further, if the act of the entity that uses fake news leads to a disposal of property contrary to the person's intention, then Article 286 of the Criminal Code applies, i.e., concerning the offense of fraud. Pursuant to Article 286 of the Criminal Code, any person who, with the intent to achieve a material benefit, causes another person to unfavorably dispose of their property, or the property of a third party by misleading the person or by taking advantage of a mistake or an inability to properly understand the action undertaken is liable to the penalty of deprivation of liberty for a term of between six months and eight years.

Another group of crimes where fake news may be used are prohibited acts committed to the harm of the state. Article 132 of the Criminal Code can serve as the best example; according to it, any person who, while providing intelligence services to the Republic of Poland, misleads a Polish state authority by delivering counterfeit or altered documents or other items, or by concealing the true information or furnishing false information of essential importance to the Republic of Poland is liable to the penalty of deprivation of liberty for a term of between one and ten years.

A count of a fake news offense can also be considered where a group of people or a person is insulted because of their ethnic or national affiliation. In this case, the provision of Article 257 of the Criminal Code applies, concerning the offense of racism. It stipulates expressly that any person who publicly insults a population group or an individual because of their national, ethnic, racial, or religious affiliation, or irreligiosity, or for these reasons violates the personal integrity of another person is liable to the penalty of deprivation of liberty for a term of up to three years. Article 256 of the Criminal Code is construed in the same way; according to it, any person who publicly promotes a fascist or other totalitarian system of state or incites hatred based on national, ethnic, racial, or religious differences, or on the grounds of irreligiosity is liable to fines or the penalty of restriction or deprivation of liberty for a term of up to two years.

If fake news is used to report a non-existent event, Article 224a of the Criminal Code applies, according to which any person who, knowing that there is no danger, reports an event that threatens the life or health of many persons or property of substantial value, or creates a situation that is meant to persuade others that such a danger exists, as a result of which a public utility institution or an authority responsible for ensuring public security, order, or health is induced to act in order to eliminate such a danger is liable to the penalty of deprivation of liberty for a term of between six months and eight years.

To complete the picture, one should mention that criminal liability for fake news is provided for in the Act of 18 December 1998 on the Institute of National

Remembrance – Commission for the Prosecution of Crimes against the Polish Nation.[88] Pursuant to Article 55 of this Act, any person who publicly and contrary to the facts denies crimes committed against the Polish nation is liable to fines or the penalty of deprivation of liberty for a term of up to three years. It is also important that, in this case, the sentence is made public.

9. Legal liability for the creation, dissemination, and use of fake news (civil liability)

Civil law is an area characterized by specific regulations that are of a different nature than, for example, criminal regulations. Civil law belongs to the domain of private law, which governs relations between autonomous subjects who have their own areas of property interests, as well as non-property (personal) interests, subject to protection under law.[89] With its specific nature, civil law regulates property and non-property relations between individuals and bodies corporate and other entities on the equal rights basis.[90] There is no room for arbitrariness on the part of public authorities to the extent present, for example, in the domain of criminal law. The peculiarity of civil law is clearly seen in the provision of Article 1 of the Civil Code,[91] according to which the Code governs civil law relations between individuals and bodies corporate.[92]

Hence, the civil law system also contains regulations that indirectly apply to the fake news phenomenon under our analysis. If the information contained in fake news violates an individual's or institution's personal interests, then it activates the entire mechanism of personal rights protection.

In Poland, the legal basis for broadly understood civil law is the Civil Code. In the context of fake news, Article 23 of the Civil Code is important. This provision stipulates that a human being's personal interests, particularly health, freedom, dignity, freedom of conscience, name or pseudonym, image, privacy of correspondence, inviolability of the dwelling, and scientific, artistic, inventive, or improvement achievements, remain under the protection of civil law, irrespective of any protection enjoyed under other regulations.

Related to the above provision is Article 24 of the Civil Code, according to which, any person whose personal interests are threatened by another person's action may demand that the action be ceased unless it is not unlawful. In the case of

88 Act of 18 December 1998 on the Institute of National Remembrance — Commission for the Prosecution of Crimes against the Polish Nation. Journal of Laws of 2021, item 177.
89 Radwański and Olejniczak, 2015, p. 2 ff.
90 Gniewek and Machnikowski (ed.), 2014, p. 4 ff.
91 Act of 23 April 1964, Civil Code, consolidated text Journal of Laws of 2017, item 459, 933 and 1132.
92 Bielski, 2015.

infringement, they may also demand that the person committing the infringement perform the actions necessary to remove its effects, including that the person make a declaration of the appropriate form and substance. Under the terms provided for in the Code, they may also demand monetary compensation, including payment of an appropriate amount to a specific public cause. If damage to property has been caused as a result of infringement, the aggrieved party may demand that the damage be remedied under the generally applicable laws.

Under these two basic provisions, a person whose good name has been damaged by the creation and spread of false or deceitful information (fake news) will have the right to bring civil action for infringement of personal interests within the Polish legal system. In this type of lawsuit, a person whose personal interests have been violated by fake news may claim *first*; property remedies to protect their personal rights, including compensation for material damage, or pecuniary compensation for non-property damage; *second*; non-property remedies to protect their personal rights, including a demand to cease such an offensive act, an action for determination (basically to demand determination of whether the personal interest has been infringed or threatened), or a demand for removal of the effects of the infringement.

10. Legal liability for the creation, dissemination, and use of fake news (administrative liability)

Administrative law is another legal domain that governs fake news-related acts. The area of administrative law is a special domain of legal regulations. Legal doctrine commonly defines public law as an area that regulates legal relations, where at least one party is a state, central or local government body, or another organization established by law to pursue the interests of the state community or narrower groups of the population (public interests).[93]

Therefore, the key concept here is the concept of authority as characteristic of public bodies, the powers of which extend far beyond the individual's legal position. In other words, while in civil law, individuals can position themselves in a binding legal relation on equal rights, under administrative law, individuals no longer have such a privilege and freedom. A public body, by virtue of its authority, conferred by the legislator, has the power to independently shape individuals' legal position.[94] Public authorities' broad range of activity results from the scope of public tasks assigned to them. Therefore, scholarly literature rightly emphasizes that the concept of public authority is exceptionally broad and

93 Korycki, Kuciński, and Trzciński et al., 2010, p. 92.
94 Cieślak (ed.), 2012, p. 19.

> is not limited to the sphere of a strictly understood public power, but covers all forms of performance of public tasks, even those where no element of power is exercised but with influence on the legal position of an individual.[95]

In this area, it is clear that there is a certain state of inequality between subjects in the domain of administrative law, which is quite different to civil law or, more broadly, to private law, as mentioned before. As a result, a public body is stronger in terms of competence, organization, and powers than an individual, which enables it to arbitrarily (authoritatively) influence an individual by influencing (shaping) the legal position of a person or other organizational units through the issuance of a decision, which is defined as one of the forms of public administrative activity.[96]

Fake news in the domain of administrative law will manifest itself, *inter alia*, in terms of decision-making powers. It may happen that information issued by a public authority based on fake news may be used as the basis for an administrative decision. This, of course, should be treated as an example only, since administrative decisions may concern an unlimited number of cases arising in various areas of administrative law. However, there is currently no legal regulation in Poland falling under the scope of administrative law that would provide for some kind of liability, for example, for issuing a decision taken as a result of fake news.

11. Legal liability for the creation, dissemination, and use of fake news (professional or corporate disciplinary liability)

Disciplinary liability is a specific type of legal liability. In dictionary terms, 'discipline' means subordination to the law regulating the internal relations of a certain community, organization, social group, or professional group.[97] Disciplinary regulations are those aimed at maintaining order within a social or professional structure. Hence, at the root of disciplinary liability is an attempt to ensure, by means of legal regulations, that the members of a professional group discharge their duties properly. The due performance of such duties is a guarantee of reliable outcomes, while adherence to professional ethos guarantees integrity in such performance.

The basic assumption is that disciplinary liability has an ordering and corrective power in relation to potential professional negligence. What comes to the fore is the preventive aspect of disciplinary liability regulations. Corporation members are subject to an *a priori* normative duty of care model in the performance of official

95 Bączyk–Rozwadowska, LEX 191278/1.
96 Knosala, 2011, p. 23.
97 Skorupka, Auderska and Łempicka, 1968, p. 147.

tasks. This requirement is complemented by moral standards, which are set and play a crucial role especially in those professional groups that are counted among public servants (public administration).

The highest measure of duty of care in a professional group member's activities and ethical requirements form the model of disciplined corporate operation. Assigned tasks determine the type of profession and make it possible to distinguish certain corporations and social groups from others.[98] Infringement upon the above models of conduct is considered an act contrary to a corporation's professional deontology. In view of the above, on the one hand, it is emphasized that disciplinary laws and procedures are a set of legal provisions governing liability for acts that violate the official duties established for a profession or social group, and the types of penalties for such acts, as well as the rules and procedures to be followed in the event of a breach of official duties.[99] On the other hand, it is noted that the formula of disciplinary liability provides for a straight-line connection between a model of due care/ professionalism in the performance of official duties with a model of an ethical and moral attitude among individuals working professionally in a specific corporation, which, apart from the substantive requirements specific for that group, abides by the values that group recognizes as key or primary. Elsewhere, it is noted that the rules of disciplinary laws are intended to raise and guarantee the prestige and ethos that distinguish a social group. This is achieved, *inter alia*, by guaranteeing the jurisdictional independence of members within institutions or corporations operating according to specific rules.[100] There is also a view that disciplinary liability is a measure that imposes self-discipline and self-control on organizationally and legally separated social groups.[101]

In view of the above, there is no doubt that liability for fake news under disciplinary laws will arise—as elsewhere under the legal liability types analyzed above—for the creation, dissemination, or both of false news. A few constituent elements are important here, including: *first*; whether the person involved in the act related to fake news belongs to a specific professional group governed by its disciplinary rules, *second*; whether the context of the fake news used is associated with the professional activities of such person; *third;* whether the content of the fraudulent message harms in any way the good name, prestige, or confidence in a member of a professional group.

All legal professions provide for this type of disciplinary liability as a general rule. *Inter alia*, the profession of an advocate (attorney) in Poland is organized under the Bar Act of 26 May 1982. [102] A special Section VIII of the Act provides for the rules of disciplinary liability. As a rule, according to Article 80 of the Bar Act, advocates

98 Wielec (ed.), 2018, p. 21.
99 Paśnik, 2000, p. 8.
100 Zubik and Wiącek, 2007, p. 70.
101 Kozielewicz, 2011, p. 85.
102 Journal of Laws of 1982, No. 16, item 124.

and advocate trainees are subject to disciplinary liability for conduct contrary to the law, ethics, or dignity of the profession or for breach of their professional duties, and for advocates, failure to comply with the obligation to conclude an insurance contract. The following disciplinary penalties apply: 1) warning; 2) reprimand; 3) financial penalty; 4) suspension from professional activities for a term of between three months and five years; 6) expulsion from the bar. Additionally, the Code of Advocate Bar Ethics and Dignity of the Profession (Code of Bar Ethics) applies.[103] According to the Code, the principles of bar ethics result from ethical standards adapted to the profession of an advocate. A violation of the dignity of the advocate profession is such conduct of an advocate as could degrade them in public opinion or undermine confidence in the profession. It is the duty of an advocate to observe ethical standards and protect the dignity of the advocate's profession. It is the duty of an advocate practicing abroad to comply with the standards contained in the Code as well as with the standards of the bar ethics in the host country.

Similar provisions are contained in the act on attorneys-at-law (legal advisers),[104] where Chapter 6 stipulates that attorneys-at-law and trainee attorneys-at-law are subject to disciplinary liability for conduct contrary to the law, ethics, or the dignity of the profession, or for breach of their professional duties. Attorneys-at-law are also subject to disciplinary liability for failure to comply with the obligation to conclude an insurance contract. The following disciplinary penalties apply here: 1) warning; 2) reprimand; 3) financial penalty; 4) suspension of the right to practice as an attorney-at-law for a term of between three months and five years, and for trainee attorneys-at-law, suspension of the rights of a trainee for a term of between one and three years; 6) deprivation of the right to practice as an attorney-at-law, and for trainee attorneys-at-law, expulsion from the traineeship.

As with advocates, attorneys-at-law are bound by the Code of Ethics of Attorneys-at-Law.[105] According to the Code, an attorney-at-law who practices the profession in a free and independent manner serves the interests of the justice system as well as those whose rights and freedoms have been entrusted to them for safeguarding. The profession of an attorney-at-law as protected under the Constitution of the Republic of Poland is one of the guarantees of respect for the law. It is a profession of public trust that respects the ideals and ethical obligations formed in the course of its

103 Announcement of the Praesidium of the Bar Council of 27 February 2018 on the publication of the consolidated text of the Code of Advocate Bar Ethics and Professional Dignity (Code of Bar Ethics). Pursuant to Resolution No. 52/2011 of the Bar Council of 19 November 2011, the consolidated text of the Collected Rules of Advocate Bar Ethics and Dignity of the Profession (the 'Code of Bar Ethics') is hereby published as adopted by Resolution of the Bar Council on 10 October 1998 (Resolution No. 2/XVIII/98), as amended by Resolution of the Bar Council No. 32/2005 of 19 November 2005, Resolutions of the Bar Council Nos. 33/2011 - 54/2011 of 19 November 2011 and Resolution of the Bar Council No. 64/2016 of 25 June 2016.

104 Attorneys-at-law Act of 6 July 1982.

105 Appendix to Resolution No. 3/2014 of the Extraordinary National Assembly of Attorneys-at-Law of 22 November 2014.

practice. Defining the rules of conduct in professional and corporate life contributes to dignity and integrity in the practice of the profession of legal adviser.

It is therefore clear that the above rules provide for disciplinary liability for the creation, use, etc., of fake news within these corporations.

12. Attempts to formulate proposals for international and national level legal regulations on mitigating or combating fake news

In Poland, there is currently a debate on the shape of the legal regulations regarding fake news analyzed in this paper. There is no doubt that the presence of fake news has become so widespread that it warrants the drafting of a single act to comprehensively regulate issues related to its creation and dissemination. There is also no doubt that, in this case, the object of regulation is extremely sensitive. The 1997 Constitution of the Republic of Poland guarantees the freedoms of communication and expression, as well as the freedom to source and disseminate information. The Constitution also stipulates that preventive censorship of the means of social communication and licensing of the press are prohibited. Therefore, the proverbial golden mean is needed between the above key constitutional guarantees and the reliability and truth of information created, held, and disseminated.

In view of the above, it should be noted that the legislative work in progress on the subject matter in Poland includes a draft act on the protection of the freedoms of social network users, which was sent on 22 January 2021 to the Chancellery of the Prime Minister with a request for entry on the list of the Council of Ministers' legislative work.[106]

For the purposes of briefly presenting this regulation, one should first note that the act sets out to establish conditions for: 1) supporting freedom of expression; 2) guaranteeing the right to truthful information; 3) improving the degree of protection of human rights and freedoms on online social networks made available in the territory of the Republic of Poland, with at least one million registered users; 4) observance by online social networks of the freedom of expression, the freedom to source and disseminate information, the freedom to express religious and philosophical views and beliefs, and the freedom of communication.

The draft act contains a glossary of statutory terms, i.e., a number of legal definitions previously mentioned in our study that are extremely important to analyzing the fake news phenomenon. The glossary provides the definitions, *inter alia*, of the following terms:

106 See: https://bit.ly/3EG1Kqy.

first, an online social network service, understood as an electronically-provided service that allows users to share any content with other users or the general public and is used by at least one million registered users in the country;

second, a service provider or the provider of online social network services consisting of the storage of user-provided information on the online social network at their request, with at least one million registered users;

third, a country representative, which is an individual or a body corporate with a place of residence or registered office in the territory of the Republic of Poland, having the exclusive right to represent the service provider in the territory of the Republic of Poland and to conduct internal audit and control procedures on its behalf;

fourth, a user, i.e., the service recipient and an individual, a body corporate, or an organizational unit without legal personality that uses the online social network service, even in the absence of a user profile;

fifth, a user profile, which is understood as settings comprising a social network service user's working environment;

sixth, disinformation, which is false or misleading information produced, presented, and disseminated for profit or violation of the public interest;

seventh, content of criminal nature, which is understood as content that praises or incites the commission of prohibited acts as specified under articles of the Criminal Code[107] or which meet the criteria of a prohibited act;

eighth, illegal content, which is content that violates personal rights, disinformation, or content of a criminal nature, or content that violates morality, particularly by disseminating or praising violence, distress, or humiliation;

ninth, a limitation of access to content, which covers all acts and omissions taken in any form with a view to limiting access to content posted on an online social network service, including by deleting user-posted content that is not illegal content, or a limitation of access to content through the algorithms used by a service provider or tags that indicate possible violations in the published content;

tenth, a limitation of access to the user's profile, which is removing or preventing access to the user's profile, limiting or preventing the sharing of content on the user's profile with other users, including through the algorithms used by the service provider, which limit the display of content posted by the user, or tags that indicate possible violations in the published content.

The most important body examining the reliability and truthfulness of information under this draft act is the Freedom of Speech Council. It will be a public administration body upholding social network services' observance of the freedom of

107 These are Articles 117–119, 127–130, 133, 134–135, 137, 140, 148–150, 189–189a, 190a, 194–204, 222–224a, 249–251, 255–258 and 343 of the Criminal Code Act of 6 June 1997.

expression, the freedom to source and disseminate information, the freedom to express religious and philosophical views and beliefs, and the freedom of communication.

The Council would be composed of a chairman in the rank of a secretary of state and four members. As member of the Council, a person may be appointed who: 1) has only Polish citizenship and enjoys full public rights; 2) has full legal capacity; 3) has not been validly convicted for an intentional crime or an intentional fiscal crime; 4) in the period from 22 July 1944 to 31 July 1990, has not worked or served in state security bodies within the meaning of Article 2 of the Act of 18 October 2006 on disclosure of information on documents of state security bodies from the years 1944 to 1990 and the content of such documents (Journal of Laws of 2020, item 2141) and has not cooperated with these bodies; 5) has an unblemished reputation; 6) has a university degree in law or the necessary knowledge in the field of linguistics or new technologies. Further, a deputy of the Sejm, senator, member of the European Parliament, councilor, community head (mayor), deputy community head (deputy mayor), community secretary, community treasury officer, district board member, district secretary, district treasury officer, province board member, province treasury officer, or province secretary may not be appointed as a member of the Council. Council members' term of office is six years. The chairman is appointed by the Sejm of the Republic of Poland by a 3/5 majority of votes with at least half of the statutory number of deputies attending. If, in the first vote concerning the appointment of the chairman, none of the candidates wins a 3/5 majority of votes, the vote is to be repeated, with the Sejm of the Republic of Poland appointing the chairman by a simple majority of votes.

The process of dealing with fake news is a staged procedure, where:

Stage 1 is the detection of information that meets the criteria of fake news, which has been disseminated via a specific social network service.

Stage 2 is the submission of a complaint to the service provider, specifying that a piece of information detected in the service is information that meets the criteria of fake news. The service has a time limit of 48 hours to consider the complaint.

Stage 3 is the response from the service that has disseminated the fake news. When considering complaints, the service has two options: either it accepts the complaint and the procedure ends at this point, or it rejects the complaint, which warrants the right to appeal the rejection decision before the Freedom of Speech Council, within a time limit of seven days. Thus, the fourth stage of the procedure opens.

Stage 4 is the activation of the Freedom of Speech Council, which, after receipt of the appeal, has a time limit of seven days to analyze it and decide. The Freedom of Speech Council may uphold the service's decision and dismiss the appeal, or it may uphold the appeal and thus, under the force of the Council's decision, have the content on the service rectified. Dismissing the appeal begins the fifth stage of the procedure, and accepting the appeal ends the procedure for the case.

Stage 5 is opened by the above-mentioned dismissal of the appeal and provides the option to bring a complaint to the Supreme Administrative Court[108] within a time limit of 30 days. In such a case, the interested parties become subject to the administrative court procedure, while the application of the act on the protection of the freedoms of social network users ends.

It should also be noted that, based on the draft act, options are being analyzed for the introduction of a 'name-blind civil action.' In short, this provision would be construed so that a person whose personal rights are violated online by another unidentified person will be able to bring action for the protection of their rights without providing the respondent's details. For action to be effectively brought before the court, it is sufficient to identify the URL where the offensive content was published, the date and time of publication, and the name of the user profile or login.

Currently, the act is at the draft stage, but the procedure laid down in it is the first and, so far, the only attempt to comprehensively regulate the issue of fake news, *inter alia*, in the Polish legal system.

108 The Polish legal system features administrative courts that administer justice through the control of the activities of public administration and the resolution of competence and jurisdiction disputes between the bodies of local government units, local government appeal boards, and Between these bodies and government administration bodies. Administrative courts comprise the Supreme Administrative Court and provincial administrative courts. Cases falling within the jurisdiction of administrative courts are examined in the first instance by provincial administrative courts. The Supreme Administrative Court supervises the activity of provincial administrative courts as regards adjudication, in the manner specified by statutes. In particular, it examines appeals to these courts' decisions and adopts resolutions clarifying legal issues. It also examines other cases within the jurisdiction of the Supreme Administrative Court under other acts. The legal basis for the organization of the Supreme Administrative Court is the Act of 25 July 2002 – Law on the System of Administrative Courts.

Bibliography

Bączyk–Rozwadowska, K. 'Glosa do wyroku SN z dnia 7 marca 2013 r., II CSK 364/12', *LEX* 191278/1.

Bąk, A. (2016) 'Serwisy społecznościowe – efekt Facebooka i nie tylko', *Media i Społeczeństwo*, 2016/6, pp. 134–146.

Bakalarski, K. (2017) 'Postprawda', *Studia Bobolanum*, 28(3), pp. 115–136.

Bąkowicz, K. (2019) 'Wprowadzenie do definicji i klasyfikacji zjawiska fake newsa', *Studia Medioznawcze*, 20(3), pp. 280–289.

Bąkowski, T. (2017) 'Kilka krytycznych uwag i spostrzeżeń dotyczących formułowania definicji legalnych we współczesnym prawodawstwie', *Gdańskie Studia Prawnicze*, Vol. XXXVI, pp. 57–67.

Bielski, P. (2015) 'Glosa do uchwały SN z dnia 5 grudnia 2014 r., III CZP 93/14', *LEX* 270731/1.

Brenda-Mańkowska, A. (2019) 'Fake news, clickbait, żebrolajki - problemy mediów w dobie postprawdy' in Tanaś, V., Welskop, W. (ed.) '*Mass media we współczesnym świecie*', Łódź: Wydawnictwo Naukowe Wyższej Szkoły Biznesu i Nauk o Zdrowiu.

Bryndal, D., Kochański, P. (1998) 'Multimedialne platformy cyfrowe', *Rzeczpospolita*, *Prawo* supplement.

Całek, A. (2013) 'Tabloidyzacja a konwergencja mediów – obraz wzajemnych zależności w polskim dyskursie publicznym. Analiza przypadku Tomasza Lisa', Środkowoeuropejskie studia polityczne, 2013/2, pp. 311–328; https://doi.org/10.14746/ssp.2013.2.17.

Cieślak, Z. (ed.) (2012) *Nauka administracji*. Warsaw: Wolters Kluwer Polska.

Czaplicka, M. (2014) *Zarz*ądzanie kryzysem w social media. Gliwice: Helion.

Czarny, E. (2004) 'Globalizacja od A do Z' [Online]. Available at: https://www.nbp.pl/publikacje/globalizacja/ (Accessed: 10 April 2021).

Daniel, K. (2018) 'Media wobec fake newsów - motywy kreowania, przyjmowane postawy, podejmowane działania', *Polityka i Społeczeństwo*, 16(3), pp. 95–114.

Dąbrowska, I. (2020) 'Deepfake, Nowy wymiar internetowej manipulacji', *Zarządzenie Mediami*, 8(2), pp. 89–101.

Delińska, L. (2018) 'Media społecznościowe jako determinanta rozwoju usług turystycznych', *Ekonomiczne Problemy Usług*, No. 130, pp. 17–25; https://doi.org/10.18276/epu.2018.130-02.

Dębowski, J. (2014) 'O klasycznej koncepcji prawdy i jej filozoficznych podstawach. Czy w Matrixie możliwa jest prawda?' in Kiklewicz, A., Starzyńska-Kościuszko, E. (ed.) '*Oblicza prawdy w filozofii, kulturze, języku*', Olsztyn: Instytut Filozofii – Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.

Donecki, D. 'Portale społecznościowe w życiu młodzieży szkolnej' [Online]: Available at: http://www.publikacje.edu.pl/pdf/11046.pdf. (Accessed 17 April 2021).

Dziwulski, J., Ogrzebacz, T. (2017) 'Media społecznościowe jako element strategii zarządzania relacjami z klientem w dobie globalizacji' in Harasim W. (ed.) '*Wpływ zglobalizowanego świata na zarządzanie*', Warsaw: Wyższa Szkoła Promocji, Mediów i Show Businessu.

Flader, M. (2018) 'Postprawda jako efekt błędnego poszukiwania prawdy', *Teologia i człowiek*, no. 4.

Flasiński, K. (2017) 'Web 2.0' in Iwasiów S., Madejski J. (ed.) in '*Interpretatywny słownik terminów kulturowych 2.0'*, Zachodniopomorskie Centrum Doskonalenia Nauczycieli and Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Gniewek, E., Machnikowski P. (ed.). (2014) ‚Kodeks cywilny. Komentarz. Wydanie 6'.

GOGOŁEK, W. (2010) 'Komunikacja sieciowa. Uwarunkowania, kategorie, paradoksy'.
JANKOWSKI, J. (2019) 'Wpływ mediów społecznościowych na kulturę i subkultury', *Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej im. Witelona w Legnicy*, No. 31, pp. 267–275.
KACZMAREK, B. (2014) 'Ekonomiczne i społeczne następstwa globalizacji', *Annales. Etyka w życiu gospodarczym*, 17(1), pp. 35–46.
KAPLAN, A. M., HAENLEIN, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons,* 53(1), pp. 59–68; https://doi.org/10.1016/j.bushor.2009.09.003.
KAZANOWSKI, D., (2010) '*Definicja social media'* [Online]. Available at: https://networkeddigital.wordpress.com/2010/04/17/definicja-social-media/, (Accessed 14 May 2021).
KORYCKI, S., Kuciński, J., Trzciński, Z., et. al. (2010) *Zarys prawa*. LexisNexis.
KOZIELEWICZ, W. (2011) 'Odpowiedzialność dyscyplinarna i karna notariusza – wzajemne relacje', *Rejent*, 21(10), pp. 82–102.
KNOSALA, E. (2011) *Zarys teorii decyzji w nauce administracji*. Warsaw: Wolters Kluwer Polska.
KREFT, J. (2020) 'Współczesne media' in Hofman, I., Kępa-Figura, D. (ed.) *'Przemoc w mediach'*, Vol. 1.
KUŹNAR, A. (2019) 'Cyfryzacja polskiej gospodarki i wykorzystanie innowacji informatycznych oraz big data przez polskie przedsiębiorstwa', in Kowalski, A.M., Weresa, M.A. (ed.) '*Polska – raport o konkurencyjności, konkurencyjność międzynarodowa w kontekście rozwoju przemysłu 4.0*' [Online], pp. 275–294. Available at: https://bit.ly/3CIcgvN.
LASKOWSKA, M. (2012) 'Komunikacja za pomocą social media – możliwości i zagrożenia. Zarys problematyki' in Biedroń M., Wawrzak-Chodaczek M. (ed.) '*Komunikacja – (po) rozumienie – obecność społeczna'*.
LASKOWSKI, J. (1991) 'Prawo naturalne', *Prawo Kanoniczne: kwartalnik prawno-historyczny*, 34(1-2), pp. 151–162.
ŁYSZCZARZ, M., SIEROCKI R., SOKOŁOWSKI M. (2018) 'Wprowadzenie. W stronę socjologii mediów. Tematy (nie)obecne', *Media – Kultura – Komunikacja Społeczna*, 14(3), pp. 5–12.
MALINOWSKI, A. (2005) 'Definicje legalne w prawie polskim', *Studia Iuridica*, Vol. XLIV, pp. 214–240.
MEJOR, J. D. (2012) 'Digitalizacja – garść refleksji', *Z badań nad książką i księgozbiorami historycznymi'*, Vol. 6, pp. 265–266.
MYŚLAK, D. A. (2019) 'Telewizja cyfrowa i jej cyfrowe pochodne a oczekiwania współczesnego odbiorcy', *Media - Kultura - Komunikacja Społeczna*, 13(1), pp. 31–57; https://doi.org/10.31648/mkks.2987.
NOWAK, P. (2019) 'Telewizja lat 90 vs telewizja XXI wieku' in Hess A., Nowina – Konopka M., Świerczyńska – Głowina, G. (ed.) '*Dynamika przemian w mediach*'.
PALCZEWSKI, M. (2017) 'Fake news. A continuation or rejection of the traditional news paradigm?', *Acta Universitatis Lodziensis Folia Litteraria Polonica*, 43(5), pp. 23–34; http://dx.doi.org/10.18778/1505-9057.43.02.
PALCZEWSKI, M. (2019) 'Dyskurs fake newsa', *Annales Universitatis Paedagogicae Cracoviensis. Studia de Cultura*, 11(1), pp. 15–31.
PAŚNIK, J. (2000) *Prawo dyscyplinarne w Polsce*. Warsaw: Muza.
Podlecki, M. (2017) 'Fake news jako zjawisko (nie tylko) medialne – część I', *BEM*, 2017/2, pp. 125–135.
RADWAŃSKI, Z., Olejniczak A. (2015) *Prawo cywilne – część ogólna*, 3rd ed. Warsaw: C.H. Beck.
SAROWSKI, Ł. (2017) 'Od internetu WEB 1.0 do internetu WEB 4.0 – Ewolucja form przestrzeni komunikacyjnych w globalnej sieci', *Rozprawy Społeczne*, 11(1), pp. 32–39.

Skorupka, S., Auderska, H., Łempicka, Z. (1968) *Mały słownik języka polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Sołtysinski, S. (ed.) *Prawo spółek kapitałowych. System Prawa Prywatnego. Tom 17B*. Warsaw: C.H. Beck.

Stecuła, K. (2017) 'Zagrożenia związane z postępem techniki na przykładzie facebooka', *Systemy Wspomagania w Inżynierii Produkcji. Jakość, Bezpieczeństwo, Środowisko*, 6(7), pp. 223–232.

Sudomir, B. (2020) 'Młode pokolenie w mediach społecznościowych' in Sawicki, A. (ed.) *Wybrane problemy zarządzania*, Pelplin: Wydawnictwo „Bernardinum".

Śledziewska, K., Włoch, R. (2020) '*Gospodarka cyfrowa, jak nowe technologie zmieniają świat*' [Online]. Available at: https://bit.ly/39wVZxk.

Tomczak, J. (2017) 'Media społecznościowe w mikroprzedsiębiorstwach', *Zeszyty Naukowe Politechniki Łódzkiej, Organizacja i Zarządzanie*, 68(1217), pp. 143–160; https://doi.org/10.34658/oiz.2017.68.143-160.

Tytko, M.M. 'Portal społecznościowy jako środek w pracy socjalnej' [Online]. Available at: https://core.ac.uk/download/pdf/53117938.pdf. (Accessed 27 April 2021).

Warzecha, K. (2017) 'Portale społecznościowe formą rozrywki i komunikacji współczesnej młodzieży – analiza statystyczna', *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, Vol. 318, pp. 84–107.

Wasiuta, O., Wasiuta, S. (2019) 'Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość', *Studia de Securitate*, 9(3), pp. 19–30; https://doi.org/10.24917/26578549.9.3.2.

Waszak, M. (2017) 'Postprawda i fake news czy weryfikacja treści i źródeł informacji?', *Refleksje*, 2017/16, pp. 173–188; https://doi.org/10.14746/r.2017.2.13.

Werner, W., Trzoss, A. (2019) 'Czy cyfrowe media wywołują społeczną paranoję? Dwugłos w sprawie obecności teorii spiskowych w mediach społecznościowych i kulturze masowej', *Nauka*, 2019/3, pp. 147–169.

Wicińska, K. (2017) 'Media społecznościowe jako narzędzie rozwijające działania marketingowe przedsiębiorstw', *Rynek – Społeczeństwo – Kultura*, 24(3), pp. 115–121.

Wielec, M. (2017) *Wartości – analiza z perspektywy osobliwości postępowania karnego*. Lublin: Wydawnictwo Academicon.

Wielec, M. (ed.) (2018) *Odpowiedzialność dyscyplinarna. Standardy polskiego systemu prawnego na przykładzie wybranych zawodów prawniczych*. Warsaw: IWS.

Wilkowski, M. (2013) *Wprowadzenie do historii cyfrowej*. Gdańsk: Instytut Kultury Miejskiej.

Wolniewicz, B. (2012) 'O pojęciu kłamstwa i zasadzie prawdomówności', *Edukacja Filozoficzna*, Vol. 54, pp. 5–27.

Woźniak-Zapór, M. (2017) 'Fake news – niebezpieczeństwo w mediach', *Bezpieczeństwo. Teoria i Praktyka*, 2017/4, pp. 99–108.

Wójcik, M. (ed.) (2019) '*Mały leksykon postprawdy - Fake news*' [Online]. Available at: https://bit.ly/3EMbe3L (Accessed: 29 April 2021).

Wyrwińska, K., Wyrwiński, M. (2018) 'Platformy internetowe jako narzędzia ekonomii współdzielenia', *Transformacje Prawa Prywatnego*, 2018/2, pp. 91–112.

Zubik, M., Wiącek, M. (2007) 'O spornych zagadnieniach z zakresu odpowiedzialności dyscyplinarnej sędziów Trybunału Konstytucyjnego – polemika', *Przegląd Sądowy*, 15(3), pp. 69–84.

Chapter II

Censorship on Digital Platforms and Social Media Versus Freedom of Expression and Pluralism: the Perspective of the Republic of Poland

Bartłomiej Oręziak

1. Introduction

It is a truism to say that the world is constantly changing, and the 21st century perhaps best exemplifies this. The process of civilization, concomitant with technological and technical progress, that we are witnessing is an expression of people's desire to simplify and expand performance and efficiency in whatever they do. This is true for the vast majority of the areas of human life, from advanced financial transactions through new ways of learning to everyday shopping or playing chess. Law—as a multifaceted plane of legal norms defining individuals' rights, freedoms, and obligations, and thus linked with most areas of human life—is no exception. This means that the legal doctrine is also involved in the debate on the practical application of advanced technologies, in terms, for example, of their potential, the definition of the rules of legal liability, and the choice of law and jurisdiction in cross-border cases. This discussion is both theoretical and practical. Legal theory deals with the conceptualization or creation of new legal frameworks or the adaptation of traditional ones that define legal norms for new technologies. On the other hand, the practice of law focuses on implementing aspects of new technologies in terms of

Bartłomiej Oręziak (2021) Censorship on Digital Platforms and Social Media Versus Freedom of Expression and Pluralism: the Perspective of the Republic of Poland. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 47–78. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_2

their actual operation and validating legal norms devised in the theory of law. Major examples of such legal considerations, if only due to the current pandemic caused by the novel coronavirus (Covid-19) virus, are digital medicine,[1] e-health,[2] m-health,[3] telehealth,[4] telemedicine,[5] telecare,[6] sensory health,[7] and medical informatics.[8]

The scholarly issue presented for analysis is of great importance, not only to the theory of law, but also for its practical value. This is because, on the one hand, the discussion about censorship on digital platforms and social media in the context of freedom of expression and pluralism from the perspective of the Republic of Poland may bring a new contribution to legal science in a dogmatic sense. On the other hand, however, this analysis will answer the fundamental question about the restriction, or even destruction, of the essence of freedom of expression and pluralism on the Internet. The real activities of the entities that control websites are on the table, as these directly affect ordinary Internet users. This undoubtedly shapes the perception of what an individual is allowed or forbidden to do, punished for, and what rights they have. This influence is even more palpable if one considers that in 2020, as many as 89% of the European Union (EU) population declared that they used the Internet,[9] which is today referred to as a globally important area of activity.[10] The popularity of the Internet as such is constantly growing, providing a platform for traditional human activity to be performed in an innovative way. This not only has positive but also negative consequences, as it results in the emergence of dangers, in the form, for example, of cybercrime, entailing criminal liability on the Internet.

In dictionary terms, censorship means "official examination, usually under governmental control, of prints, press, literary works and motion pictures, etc., exercised by a specially appointed authority that evaluates them in political or moral terms"[11] or "official control of publications, theatrical performances, radio programs, etc., [and the] evaluation [of] these in political or moral terms."[12] These definitions are, of course, correct. However, they mainly result from the negative historical connotations of censorship in Poland. These date back to 1949–1989 when Poland was under communist rule based on the censorship of public life and specifically under the Decree of 5 July 1946, which established the Main Office for the Control of the

1 Lupton, 2013, p. 257; Elenko, Underwood, and Zohar, 2015, pp. 456–461; André, 2019, p. 4.
2 Terry, 2000, pp. 605–607; Mars and Scott, 2010, pp. 237–243; de Pietro and Francetic, 2018, pp. 69–74.
3 Paglialonga et al., 2019, p. 6; Istepanian, Laxminarayan, and Pattichis, 2006, p. xxiii; Sezgin, 2018, p. 1.
4 Maimone et al., 2012, pp. 791–793; Wang et al., 2014, pp. 314–324.
5 Klar and Pelikan, 2011, p. 1119; Linkous, 2001, p. 226.
6 Držanič et al., 2019, p. 252; Afsarmanesh, Masís, and Hertzberger, 2004, pp. 211–212.
7 Gao et al., 2020, pp. 55–56.
8 Huang, 2009, p. 1423.
9 Eurostat (25 May 2021), https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ifp_iu/default/table?lang=en.
10 Dutton, 2013, p. 1.
11 Doroszewski [Online], https://sjp.pwn.pl/doroszewski/cenzura;5416093.html.
12 *Słownik Języka Polskiego* [Online], https://sjp.pwn.pl/sjp/cenzura;2447537.html.

Press, Publications, and Public Performances.[13] The decree introduced political and preventive censorship in almost every aspect of life. In today's world, however, the concept of censorship may have a slightly different meaning. It does not always mean official or state-run control, especially as far as digital platforms and social media are concerned. This digitized environment should *ex definitione* warrant a tailored meaning of censorship that is more colored with digital elements. It seems that in this case, censorship can be understood in the paradigm presented above, yet with the necessary changes. In view of this, for the purposes of this chapter, censorship will mean the control and restriction of Internet users' activities, including their publication of content, in particular on political, moral, or legal grounds. This is a key point, as in the operating practice of digital platforms and social media, it may turn out that a legal norm will not form the basis for determining the extent of rights, freedoms, and penalties.[14] It would then be nothing more than unlawful censorship, that is, a measure without a basis in the generally applicable law in Poland. On the other hand, it might well be that in a formal sense, the source of restrictions on human rights and freedoms on the Internet is in the law. In this case, the term 'lawful censorship' should be used. The latter, however, does not mean that it is reasonable or justified, but that it is based on statutory law in Poland.[15]

It therefore seems appropriate to analyze the three main research areas related to the issue delimited in the title. First, we will discuss the lawful censorship of content posted on the Internet using the example of copyright law, the original legal basis of which is found in the EU regulations within the meaning in line with the Court of Justice of the European Union's (CJEU) latest case law. Without jumping to conclusions on this point, one may safely indicate that the matter is about weighing certain legally protected interests. On the one hand, these are human rights, i.e., freedom of expression and pluralism, and on the other, the copyright holder's equitable claims. Second, no less important is an analysis of the issues of unlawful censorship on digital platforms and social media in Poland, based on the Polish legislator's latest legislative effort. Third, the first and the second areas above should provide a basis for the assessment of the consistency of lawful and unlawful censorship with the standard of freedom of expression and pluralism, as understood in accordance with the Polish system of human rights protection.

13 Decree of 5 July 1946 establishing the Main Office for the Control of the Press, Publications, and Public Performances (Journal of Laws of 1946, No. 34, item 210). Declared repealed on the basis of the Press Law Act of 26 January 1984 (consolidated text: Journal of Laws of 2018, item 1914).

14 For instance, in the cases of the numerous website rules that have been issued and are currently in operation.

15 Legal acts enacted in Poland enjoy the presumption of constitutionality until rebuttal. This means that law made, passed, and applied in the Polish legal system (in the form of an act/statute, as discussed below), including a law restricting selected human rights and freedoms, is treated as constitutional until the Polish Constitutional Court declares it inconsistent with the Constitution. In this context, see Radziewicz, 2008, pp. 55–86; and the following judgements of the Constitutional Court: 18 March 2004, case ref. P21/02; 25 May 2016, case ref. Kp 2/15; 5 May 2011, case ref. P 110/08; July 2012, case ref. K 8/10.

2. Lawful censorship in Poland in light of EU copyright

Poland, as an EU member state, is required to apply and abide by EU law and comply with the CJEU's rulings. This is an indisputable fact resulting from Poland's commitments and obligations under international agreements and treaties.[16] Another undeniable fact is that for EU law to be construed properly, CJEU rulings, which, through the court's autonomous interpretation either creates or clarifies legal norms adopted by the EU legislator, are of key importance.[17] It should be noted that the EU case law system, unlike the Polish legal system, is based on precedent and thus has a law-making character.[18] This often leads to peculiar situations, where a single provision, or even part of it, serves to build an autonomous line of case law that fundamentally changes the existing legal position on a subject. In practice, these changes may be so far reaching that questions emerge about their compatibility with the EU legislator's original intent, which represents the EU member states' intent. Due to the law-making nature of CJEU case law as emphasized above, and the obligation on the part of EU member states to apply EU law in line with the CJEU's autonomous interpretation,[19] this issue is of great significance for legal transactions.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in information society[20] (Directive 2001/29/EC), based on today's Articles 53, 62, and 114 of the Treaty on the Functioning of the European Union (TFEU),[21] concerns the legal protection of copyright and related rights within the EU internal market, with particular emphasis on information society. Pursuant to Recital 1 of Directive 2001/29/EC, its purpose is to contribute to the achievement of the objectives of the EU internal market, including the institution of a system ensuring undistorted competition.[22]

The issue of Internet censorship is related to the admissibility of content publication on the Internet in accordance with EU copyright law under Directive 2001/29/EC. In this respect, proper interpretation of Article 3(1) of Directive 2001/29/EC is key. It states that:

> Member States shall provide authors with the exclusive right to authorize or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access them from a place and at a time individually chosen by them.

16 Treaty of Accession between the Kingdom of Belgium … the Slovak Republic, concerning the accession of the Czech Republic … the Slovak Republic to the European Union, signed in Athens on 16 April 2003, OJ L 236, 23.9.2003 (Journal of Laws of 2004, No. 90, item 864).
17 Helios and Jedlecka, 2018, pp. 134–141.
18 Zawidzka-Łojek and Grzeszczak, 2015, pp. 3–5.
19 Helios and Jedlecka, 2018, pp. 134–141.
20 OJ L 167, 22.6.2001, p. 10.
21 OJ C 326, 26.10.2012, p. 1.
22 Rosenmeier, Szkalej, and Wolk, 2019, p. 216.

As it has turned out in the practice of the application of Directive 2001/29/EC, the concept requiring the use of advanced interpretation techniques has been (and still is) 'communication available to the public,' as used in Article 3(1) of that directive. It seems that at least two scholarly problems are inescapable with this point. First, does the inclusion of a hyperlink to a protected work that is freely accessible on another website without the copyright holder's permission constitute 'communication to the public' within the meaning of that provision? The facts associated with this problem carry, in principle, a negative load. An example may be a case where computer software, which is freely available on a website, is made available on another website without the developer's license. Another typical example would be downloading videos or songs and sharing these instantaneously without their creators' consent. Second, does the inclusion of a hyperlink to a protected work that is freely accessible on another website with the copyright holder's permission constitute 'communication to the public' within the meaning of the same provision? The facts associated with this problem do not, in principle, carry a negative load. An instance of this may be downloading a photo that is freely available on a website with the photographer's consent and then making it available on another website, with attribution of the photo's source. Another typical example is the use of the 'share' function that most digital platforms and social media offer.

As is clear, the difference between the first and the second problem is whether or not the copyright holder has given their consent. This is of key importance from the interpretation point of view, because, as it will turn out, this factor is a root cause of the CJEU's adoption of a different autonomous interpretation. The reconstruction of the legal model for the admissibility of hyperlinking on the Internet from Poland's perspective as an EU member state will only follow once the proper understanding of the law with respect to these two issues has been established. The CJEU's case law, which brings a novel normative solution to the Polish legal system, is gaining importance.

Poland has transposed a number of legal measures that have implemented not only Article 3(1) of Directive 2001/29/EC but generally all the legal norms contained in this directive.[23] The transposition was made under the Copyright and Related Rights Act of 4 February 1994.[24] However, Polish implementation was based on the literal construction of these provisions, established as at 22 May 2001, in fulfilment of the obligation under Article 13 of Directive 2001/29/EC.[25] This means that

23 Act of 1 April 2004 amending the Copyright and Related Rights Act (Journal of Laws of 2004, No. 91, item 869).

24 Copyright and Related Rights Act of 4 February 1994 (consolidated text: Journal of Laws of 2020, item 288).

25 Pursuant to Article 13 of Directive 2001/29/EC, EU member states were obliged to bring into force the laws, regulations, and administrative provisions necessary to comply with the directive at the latest by 22 December 2002. Poland, however, joined as an EU Member State on 1 May 2004 pursuant to the Accession Treaty. Therefore, Directive 2001/29/EC was transposed into the Polish normative system, justifiably, at a later date.

Poland has demonstrated the best faith in the objectives of that act of EU law and, in general, in the EU legislation imposing obligations on it. One must not forget that the literal construction of Directive 2001/29/EC can and should be deemed equivalent to the EU member states' intent, as represented by the European Parliament and other EU institutions. No other conclusion may be made than that the standards of the EU copyright law agreed by the EU member states have the wording established as at 22 May 2001 and the form of an EU directive. An EU regulation was not issued back then, and rightly so, probably to avoid the direct effect of EU law within national legal systems. EU member states wanted to remain autonomous as to the measures implementing the normatively defined objective of Directive 2001/29/EC. However, in the context of its Article 3(1), it seemed, at that moment, sufficient to transpose the legal norm and understand it in accordance with the literal rule of interpretation, where the concept of communication to the public has its common and ordinary meaning. That was the case until the GS Media ruling.[26] At that moment, the CJEU obtruded with its powers of autonomous construction. In its rulings in the cases of GS Media, FilmSpeler,[27] The Pirate Bay,[28] and Renckhoff,[29] the CJEU gave autonomous interpretations of Article 3(1) of Directive 2001/29/EC that drastically changed the understanding of the concept of communication to the public, based on the argument that the concept has a specific meaning in EU law.

In view of the binding nature of the CJEU's autonomous interpretation,[30] this is tantamount to the introduction into the legal system of a specific type of construction norm that determines the actual content of Article 3(1) of Directive 2001/29/EC. It radically changes the rules of the admissibility of hyperlinking on the Internet from the copyright perspective, a matter that has thus far been understood otherwise. As a result, Poland was obliged to construe its legal measures implementing Directive 2001/29/EC through the prism of the CJEU's autonomous interpretation in the context of its Article 3(1). This is a very interesting legal problem. One may conclude that to issue an EU directive does not clearly mean to avoid the risk of the unification of legal norms under a fixed EU standard. Such a risk may be realized against the intent of an EU member state which, through its representatives in the European Parliament, has agreed to adopt a legal act with the rank of a directive. It is sufficient for the CJEU to issue an autonomous interpretation that is directly applicable and gives rise to an EU member state's obligation to construe its legal system specifically

26 Judgement of the Court of Justice of the European Union of 8 September 2016 in case C-160/15 GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker (ECLI:EU:C:2016:644).

27 Judgement of the Court of Justice of the European Union of 26 April 2017 in case C-527/15 Stichting Brein v Jack Frederik Wullems, also trading under the name 'Filmspeler' (ECLI:EU:C:2017:300).

28 Judgement of the Court of Justice of the European Union of 14 June 2017 in case C-610/15 Stichting Brein v Ziggo BV and XS4ALL Internet BV (ECLI:EU:C: 2017:456).

29 Judgement of the Court of Justice of the European Union of 7 August 2018 in case C-161/17 Land Nordrhein-Westfalen v Dirk Renckhoff (ECLI:EU:C:2018:634).

30 Helios and Jedlecka, 2018, pp. 134–141.

in accordance with the proposed standard. It seems reasonable to conclude that the CJEU's case law may lead to the alteration of the normatively envisaged objectives of an EU directive. This is because the CJEU's autonomous interpretation imperatively and directly unifies the understanding of a directive's legal norms and thus indirectly influences the application of all legal measures taken to bring it into force in the national legal system. This is precisely the case with Article 3(1) of Directive 2001/29/EC. The conclusion is, therefore, that the autonomous interpretation the CJEU issued actually made it a legal norm of an EU regulation. In simplified terms, it turns out that the CJEU has the powers to change the legal nature of an EU legal norm from indirectly to directly applicable.

Therefore, Directive 2001/29/EC could be of significance for the Polish legal system in at least two dimensions. The first one would be typical and consistent with the essence of a legal act of this rank. The other one, in turn, would be burdened with the CJEU's autonomous interpretation. This could be because in the CJEU's ruling in the Svensson case, it decided that harmonization under Directive 2001/29/EC is exhaustive and is achieved when various national laws governing an issue are replaced by a single EU standard.[31] Therefore, from the Polish perspective, it makes no difference that Directive 2001/29/EC was passed as an EU directive and not an EU regulation. The CJEU could do so from the legal point of view, and it did. In simplified terms, it changed the legal rank of Directive 2001/29/EC, of course in observance of the transposition procedure into national law and therefore national autonomy as to the choice of measures to achieve the objectives, but without options to shape the interpretations. Whether the reader agrees with the above view or not, it is an undeniable fact that, in accordance with EU primary law, Poland is obliged under international law to construe its national law in line with the CJEU's autonomous interpretation of Article 3(1) of Directive 2001/29/EC.

From the point of view of the construction of Article 3(1) of Directive 2001/29/EC as regards restrictions on posting content on the Internet due to copyright, the key is the autonomous interpretation made in the ruling in the GS Media case, where, for the first time, the CJEU introduced clear limits on the admissibility of hyperlinking on the Internet. The CJEU decided that Article 3(1) of Directive 2001/29/EC should be construed so that in order to establish whether placement on a website of hyperlinks to protected works that are freely accessible on another website without the copyright holder's permission constitutes 'communication to the public' within the meaning of that provision requires a determination as to whether such hyperlinks were made available for non-commercial purposes by a person who did not know or could not reasonably know about the unlawful nature of the publication of these works on that other website or whether, on the contrary, such hyperlinks were made available for commercial purposes, in which case such knowledge is to be presumed.[32] The CJEU has consistently upheld that position in subsequent cases. First,

31 Zawidzka-Łojek and Grzeszczak, 2015, pp. 3–5.

32 Rosati, 2017, pp. 1221–1230; Radosavljev, 2017, p. 5; Long, 2018, pp. 430–433.

it issued a ruling in the Filmspeler case where, based on the rules and constituent elements set out in GS Media case, it construed the third-party liability of a dealer selling a media player with pre-installed plug-ins (otherwise available on the Internet) containing hyperlinks to freely accessible websites featuring copyrighted works that had been placed there without the copyright holders' permission (hyperlinking in physical facilities).[33] Second, it issued a ruling in The Pirate Bay case in which it made an extensive interpretation of the rules and constituent elements from the GS Media case and proposed a new principal accessory approach to hyperlinking. In that case, the CJEU decided that an Internet user justifiably incurs third-party liability in the construction of indirect communication (sharing) or aiding and abetting hyperlinking based on the *sine qua non* condition whereby if some users did not grant access to their website and did not manage it, other users would not be able to enjoy the protected work or would be able to do so only with difficulty.[34] Third, it issued a ruling in the Renckhoff case that represents a different approach to the construction of Article 3(1) of Directive 2001/29/EC, though it does not entail a departure from the rules and constituent elements given in the GS Media case ruling, which remain intact. The ruling in the Renckhoff case is specific in that it concerns a situation where a hyperlink is placed to a protected work that is freely accessible on another website with the copyright holder's permission. Therefore, the adjudicating panel proposed that Article 3(1) of Directive 2001/29/EC in the context of hyperlinking to content that is legally available on the Internet should be examined only through the prism of a single element set out in in the GS Media case judgement, while withholding the other ones, including the commercial purpose of communication. That element is the concept of the public, as contrasted with the concept of a new public, from which, for example, it follows that the concept of communication to the public within the meaning of Article 3(1) of Directive 2001/29/EC should be construed so as to include the posting on a website of a photograph that has previously been published on another website without restrictions preventing its download and with the copyright holder's consent.[35] This is a new autonomous CJEU construction for the communication of content lawfully available on the Internet, i.e., with the copyright holder's consent.[36]

The abovementioned autonomous interpretations of Article 3(1) of Directive 2001/29/EC are only examples selected for the author's analysis. This means that there are currently more such constructions.[37] Nevertheless, the abovementioned

33 Colangelo, Maggiolino, 2018, pp. 142–159; Ginsburg, 2017, pp. 4–5.

34 Visser, 2018, pp. 1025–1026; Koo, 2018, pp. 542–551; Nordemann, 2018, pp. 744–756.

35 Fernández-Díez, 2018, p. 2; Visser, 2018, pp. 183–190; Wang, 2018, pp. 61–65.

36 The considerations presented in this paragraph are contained in: Oręziak, 2017, pp. 243–253; Oręziak, 2018, pp. 199–219; Oręziak, 2018, pp. 137–157; Oręziak, 2019, pp. 181–192; Oręziak, 2019, pp. 432–448.

37 For example, the autonomous interpretation contained in: Judgement of the Court of Justice of the European Union of 9 March 2021 in case C-392/19 VG Bild-Kunst v Stiftung Preußischer Kulturbesitz (ECLI:EU:C:2021:181).

autonomous interpretations are sufficient to present a picture of the system of the admissibility of posting content on the Internet under the EU copyright law that Poland is obliged to apply. The jurisprudence of common courts in Poland confirms that this obligation is being fulfilled, with examples including the Court of Appeal in Warsaw's 21 June 2016 decision,[38] the Judgement of the District Court in Olsztyn on 6 June 2017,[39] the Judgement of the Court of Appeal in Szczecin on 24 November 2016,[40] and the Judgement of the Supreme Court on 9 August 2019.[41] These examples clearly demonstrate that Polish courts apply the CJEU's arguments and use them as the basis for decisions on legal liability for copyright infringement on the Internet. In this context, it should be emphasized that in Poland, as a rule, courts' jurisprudence does not enjoy law-making power, so court decisions are only binding on the parties to the proceedings and do not have an *erga omnes* nature.[42] Nevertheless, other subjects tend to be inspired by such decisions, which they use as a basis for their rationale in claiming that their situation is identical or substantially identical to that which has been delimited by the facts of a decided court case. Therefore, on the one hand, Polish courts decide on the basis of the CJEU's constructions, and on the other, website administrators remove content on the basis of relevant jurisprudence in cases other than those decided within such jurisprudence. In such cases, the decision makers are not Polish courts but rather website administrators who, in their reliance on such premises, must be aware that their conduct possibly borders on unlawful censorship. Although one may say that such conduct is founded on generally applicable law in Poland, it should be noted that a website administrator is not an entity with powers to apply, enforce, and interpret the law in new cases that have not yet been examined in this regard by the competent bodies established for this purpose in Poland. Thus, even where such administrators observe the duty of utmost care, a court may still decide otherwise in a new case, on the grounds, for example, of the specific nature of the facts given in the case.

In conclusion, the example of censorship based on the EU copyright law in Poland as outlined above constitutes lawful censorship, where there is undoubtedly the exercise of control over and the restriction of Internet users' activities, including their publication of content, for legal reasons. It should be emphasized that lawful censorship only means that this type of censorship is sanctioned by generally applicable law in Poland. Assessment of this issue is rather difficult, as it is about a law that

38 Decision of the Court of Appeal in Warsaw of 21 June 2016, case ref. IA Cz 723/16.

39 Judgement of the District Court in Olsztyn of 6 June 2017, case ref. VII K 5/16.

40 Judgement of the Court of Appeal in Szczecin of 24 November 2016, case ref. I ACa 1159/15.

41 Judgement of the Supreme Court of 9 August 2019, case ref. II CSK 7/18.

42 This principle stems from Article 87 of the Constitution of the Republic of Poland, defining the sources of universally binding law in Poland. For example, a specific determination of this principle is the provision of Article 365 of the Polish Code of Civil Procedure (Code of Civil Procedure Act of 17 November 1964 (consolidated text Journal of Laws of 2021, item 11). Pursuant to this provision, a final decision is binding not only on the parties and the court that has passed it, but also on other courts and other state authorities and public administration bodies, and in the cases provided for in the Act and also on other persons.

was lawfully passed in the EU and has been correctly implemented and brought into force in the Polish legal system then modified by an autonomous interpretation made by the CJEU. The pattern seems to be correct, as EU member states have agreed to it by deciding to join the EU or by ratifying the amendments to EU primary legislation, yet an interesting question is whether they were fully aware of the consequences. The brief analysis given above demonstrates that these consequences can be far reaching. In fact, it is not about the pattern itself, as it has its basis in international law, but rather about the content hidden in this pattern, and even more importantly, whether the content is consistent with freedom of expression and pluralism on the Internet, the standards of which are guaranteed in the Constitution of the Republic of Poland (Polish Constitution).[43] It is a good point of reference because the act represents the supreme and unchallenged law in our country.

3. Unlawful censorship in Poland and a new legislative initiative

As noted above, unlawful censorship in Poland carries a negative load due to its roots in the communist rule. It should therefore come as no surprise that any form of censorship on digital platforms and social media is met with a negative response from the public. This is more intense where a basis for such restrictions is comprised not of a legal norm but other types of rules, for example, the terms of use of a specific social networking or other website. Of course, an Internet user will often accept such terms of use unmindfully, but this does not mean that censorship is then warranted to have a basis in generally applicable law in Poland. This is because, pursuant to Article 87 of the Polish Constitution, the sources of universally binding law in Poland are the Constitution, statutes, ratified international agreements and regulations, and enactments of local law issued by the operation of organs in the territory of the organ issuing such enactments.[44] These legal acts may only be passed by law-making bodies that have been constitutionally empowered to do so. In turn, according to Article 91(3) of the Polish Constitution, if an agreement ratified by the Republic of Poland and establishing an international organization so provides, the laws it establishes apply directly and have precedence in the event of a conflict of laws.[45] Terms

43 The Constitution of the Republic of Poland of 2 April 1997 (Journal of Laws of 1997, No. 78, item 483; and the amendments: Journal of Laws of 2001, No. 28, item 319; of 2006, No. 200, item 1471; of 2009, No. 114, item 946).

44 See the following judgements of the Constitutional Court: of 29 November 2017, case ref. P 9/15; of 6 December 2016, case ref. SK 7/15; of 10 February 2015, case ref. SK 50/13; of 12 December 2011, case ref. P 1/11; of 10 May 2005, case ref. SK 40/02; of 7 July 2003, case ref. SK 38/01; of 10 June 2003, case ref. SK 37/02.

45 See Judgement of the Constitutional Court of 12 December 2011, case ref. P 1/11.

of use established for the operational purposes of digital platforms and social media certainly have nothing to do with generally applicable law in Poland. As such, they cannot autonomously lay down rights, obligations, or penalties, nor limit an individual's legally created rights. These documents can be issued, for example, in order to properly organize online communities, but if they introduce any restrictions on their users' rights and freedoms, they must comply with protection standards in this regard, for example, the standard of freedom of expression or pluralism that results from generally applicable law in Poland.

As the observation of the terms of use prevailing on the Internet demonstrates, this is generally not the case. The fact is that the current models of the control and restriction of Internet users' activities, including their publication of content, especially on political or moral grounds, arise from documents that have the character of internal rules. In principle, these documents make no reference to the Polish legal order, in particular due to the transnational nature of the activities of digital platforms or social media. This feature, however, cannot justify restrictions on the freedoms and rights of individuals on the Internet as provided for in the Polish normative system. All of this provides a typical example of unlawful censorship on the Internet. After all, an action contrary to such terms of use results in the removal of a post or temporary or permanent account suspension. On the other hand, if the user believes that the restriction they have suffered is unjustified, they may bring action before the courts to assert their rights and freedoms. For this reason, most terms of use specify the jurisdiction and governing law should a dispute arise.[46]

An analysis of this aspect can be applied to Facebook's terms of service,[47] which the portal has officially translated into Polish and thus have been prepared for, apply to, and are accepted by Polish Internet users. This example is relevant and, as it seems, representative, since the majority of Polish citizens use this social network service.[48] Incidentally, one can mention that the content of the terms is most likely the same for all Facebook users around the world. Section 3 of Facebook's terms of service sets out 'Your commitments to Facebook and our community.' For example, in accordance with item 2 of this section, 'What you can share and do on Facebook,' one of Facebook's objectives is to ensure that its users express themselves and share content that is important to them, but not at the expense of others' safety and well-being or the integrity of the Facebook community. For this reason, Facebook users may not use Facebook products to do or share anything that (a) breaches Facebook's terms, community standards, and other terms and policies that apply to use

46 In the context of special jurisdiction in cases of infringement of personal rights on the Internet, see: Judgement of the Court of Justice of the European Union of 17 October 2017 in case C-194/16 Bolagsupplysningen OÜ and Ingrid Ilsjan v Svensk Handel AB (ECLI:EU:C:2017:766); Judgement of the Court of Justice of the European Union of 25 October 2011 in joined cases C-509/09 and C-161/10 eDate Advertising GmbH and others v X and Société MGN LIMITED (ECLI:EU:C:2011:685).

47 *Facebook Terms of Service*, https://www.facebook.com/legal/terms.

48 According to the 'NapoleonCat' report of January 2021, Facebook is used by 57.2% of Poles (21,680,000 users).

of Facebook; (b) is unlawful, misleading, discriminatory or fraudulent; or (c) infringes or breaches someone else's rights, including their intellectual property rights. In view of the above, Facebook can remove or block content when a Polish user breaches these provisions of the terms of service. Additionally, regardless of the above, Facebook may remove or restrict access to any user content, service, or information, including a user's account, if it deems such an action reasonably necessary to avoid or mitigate adverse legal or regulatory impacts to Facebook. In practice, this means that Facebook's terms of service authorize it to remove any content that it deems problematic.

In this context, inspired by Facebook's terms of service, the discussed issue can be considered in at least two categories from the perspective of the Polish normative system. The first one is where a social network service administrator censors content on legal grounds. This is, as noted before, lawful censorship, as it is based on generally applicable law in Poland. An example of this is the censorship of content on the Internet based on copyright, as presented above. The other category is where a social network service administrator censors content, especially on political or moral grounds, but on a basis different than generally applicable law in Poland. Such other basis often comes from the specific terms of a social network site's community. This is an example of unlawful censorship, the scale of which is much larger. The rules determining its application are not always clearly defined, or rather, they are treated as a specific empowerment of the website's owner, quite apart from the human rights system recognized in Poland and from the rights and freedoms that every individual enjoys regardless of where they perform their activities.

In view of the above, coupled with the fact that in Poland, there is currently no legal framework directly concerning either the legal status of digital platforms and social media or the sanctioning of unlawful censorship, one can conclude that this issue remains in a legal vacuum. In recognition of this, the Polish Ministry of Justice, on 22 January 2021, decided to apply to the Chancellery of the Prime Minister for the entry of a draft act on the protection of freedom of speech on social network sites[49] (draft act on freedom of speech) on the list of the Council of Ministers' legislative works. This new legislative initiative appears to be a direct response to the existing regulatory gap. The message seems clear: Freedom of expression, speech, or pluralism is to be guaranteed regardless of the environment in which an individual is active, especially in the digital environment. Further, any restrictions on Polish Internet users' rights and freedoms must be normatively sanctioned.

According to the recitals of the draft act on freedom of speech, it is to be passed due to the special constitutional value of freedom of speech and with a view to strengthening this freedom's role in the search for truth, the functioning of a

49 Draft act on the protection of freedom of speech on social network sites (15 January 2021), https://www.gov.pl/web/sprawiedliwosc/zachecamy-do-zapoznania-sie-z-projektem-ustawy-o-ochronie-wolnosci-uzytkownikow-serwisow-spolecznosciowych.

democratic state, and respect for the principle of freedom of expression and human dignity. The Polish Ministry of Justice has pointed out that under the new legislation, social network sites will not be able to remove posts or block Polish users' accounts if the content posted does not violate Polish law.[50] This seems to be a basic rule that will entail severe administrative penalties if breached. The proposed legal act also aims to fill the above-indicated legal gap. The Polish Ministry of Justice has clearly confirmed this by stating that website administrators currently make independent decisions to remove posts and block user accounts, for which there is no effective remedy or appeal available, even if the user proves that they have not violated any law and that the website's action violates freedom of expression.[51] Another important objective of the draft act on freedom of speech is to ensure the effective right to truthful information and protect personal rights or interests infringed upon by anonymous Internet users.[52]

Pursuant to Article 1 of the draft act on freedom of speech, the objective of the document is, *inter alia*, to create conditions for supporting freedom of expression, improve the level of protection of human rights and freedoms on social network services made available in the territory of Poland with at least one million registered users, and uphold social network services' observance of the freedom of expression, the freedom to source and disseminate information, the freedom to express religious and philosophical views and beliefs, and the freedom of communication. The draft act on freedom of speech has a number of provisions aimed at achieving this goal. First, it defines the rules for controlling activities involving the provision of electronic services[53] via online social network sites with at least one million registered users to the extent that public authorities are in a position to guarantee their users the right to freedom of expression. Second, it defines service providers' obligations as regards guarantees of the right to freedom of expression. Third, it lays down rules for internal audit and check processes to be performed by service providers with regard to user complaints concerning violations of the right to freedom of expression. Fourth, it contains provisions on proceedings before public administration

50 Ibid.

51 Ibid.

52 This issue will be discussed in a separate chapter of this monograph by Dr. hab. Marcin Wielec.

53 Pursuant to Article 2(4) of the Electronic Services Act of 18 July 2002 (consolidated text: Journal of Laws of 2020, item 344), provision of services electronically means rendering a service provided without simultaneous presence of the parties (remotely), by transmitting data at the individual request of the service recipient, as submitted and received by means of electronic processing devices, including digital compression and data storage, which are entirely issued, received or transmitted over an information and communications technology (ICT) network within the meaning of the Telecommunications Law Act of 16 July 2004 (consolidated text: Journal of Laws of 2021, item 576). In turn, pursuant to Article 2(35) thereof, an ICT network means transmission systems and switching or diverting devices, as well as other resources, including inactive network components, which enable the broadcast, reception, or transmission of signals by means of wired, radio, optic, or other means using electromagnetic energy, regardless of the kind.

authorities and court actions in the event of the restriction of access to an electronic service provided via an online social network site.

The draft act on freedom of speech also intends to introduce a number of legal definitions into the Polish legal order. From the perspective of this study's subject matter, special attention should be paid to four definitions. The Polish Ministry of Justice proposed that the concept of an online social network service should be understood as an electronically provided service that allows users to share any content with other users or the general public and is used by at least one million registered users in the country. The definition of a user profile is also interesting, as it is intended to denote settings for social network service users' working environment. Currently, there are no such definitions in the Polish legal system. Nevertheless, from the point of view of the issue of unlawful censorship on digital platforms and social media, two other definitions are key, namely the legal definition of a limitation of access to content and a limitation of access to the user's profile. According to the draft act on freedom of speech, a limitation of access to content should be understood to cover all acts and omissions taken in any form with a view to limiting access to content posted on an online social network service, including by deleting user-posted content that is not illegal,[54] and a limitation of access to content through the algorithms the service provider uses or tags, which indicate possible violations in the published content. Further, a limitation of access to the user's profile is meant to remove or prevent access to the user's profile, limiting or preventing content sharing with other users, including through the algorithms the service provider uses to limit the display of user-posted content or tags that indicate possible violations in the published content.

The above clearly demonstrates that the proposed regulation's material scope is extremely broad. In fact, these regulations are to be applied in practice to the functionalities of a substantial part of the Internet. It seems that, for this reason, the appointment of an appropriate public administration body is an inevitable consequence of adopting the legal regulations presented here. Therefore, the draft act on freedom of speech provides for the appointment of the Freedom of Speech Council (hereinafter 'the Council'), which is tasked with upholding social network services' observance of the freedom of expression, the freedom to source and disseminate information, the freedom to express religious and philosophical views and beliefs, and the freedom of communication. It is intended to be a key public administration body that can be understood as a control mechanism for ensuring that the proposed law is applied in practice. The Council is to have a six-year term of office, and its five-person panel is to comprise experts in the field of law and new media elected by the Sejm of the Republic of Poland by a 3/5 majority of votes, which essentially guarantees a

54 According to Article 3(8) of the draft act on freedom of speech, the term 'illegal content' is understood as content that violates personal rights, disinformation, content of criminal nature, as well as content that violates morality, especially by disseminating or praising violence, distress, or humiliation.

multi-partisan and pluralist composition.[55] Meetings of the Council are to be closed to the public, and the Council is to issue its decisions, rulings, and resolutions by a majority of votes by show of hands (open vote) with three members, including the chairman, in attendance. In the event of an equal number of votes, the chairman has the casting vote. The Council's decisions and rulings are to be final and valid.

The draft act on freedom of speech clearly emphasizes that a service provider[56] is required to perform the obligations set out in the act, which is to become a universally binding law in Poland. First, it was decided that a service provider who receives over 100 user complaints in a calendar year regarding, *inter alia*, limited access to content or to users' profiles is obliged to prepare a semi-annual report in the Polish language concerning the resolution of such complaints and publish it on its social network site no later than one month after the end of a six-month period. Such reports published on an online social network site must be clearly visible and directly and permanently accessible. Regardless of this, such a service provider is obliged to file a request with the president of the Office of Electronic Communications[57] (UKE) for the publication of its report in the Official Journal of the Office of Electronic Communications no later than one month after the end of a six-month period. Second, a service provider is required to appoint at least one, but not more than three, country representatives. The representatives' responsibilities would include representing the service provider in all judicial and extrajudicial actions, considering complaints during internal audit and check processes, providing responses and any information to institutions and authorities in relation to any conducted proceedings, and participating in training organized by the president of the UKE on the current legal status regarding complaints considered during internal audit and check processes. In view of the above, the draft act on freedom of speech assumes that a service provider is obliged to immediately inform the president of the UKE about the appointment or any changes to a country representative and provide their data, including their e-mail address and address for service. Where the country representative is a body corporate,[58] a service provider provides the data of natural persons[59] authorized to act on behalf of that body corporate. Service

55 https://www.gov.pl/web/sprawiedliwosc/zachecamy-do-zapoznania-sie-z-projektem-ustawy-o-ochronie-wolnosci-uzytkownikow-serwisow-spolecznosciowych.

56 According to Article 3(2) of the draft act on freedom of speech, the term 'service provider' means a provider of online social network services, consisting of the storage of information provided by the user on the online social network upon the user's their request and with at least one million registered users.

57 Pursuant to Article 3 of the Act of 29 December 2005 on transformations and changes in the division of responsibilities and powers of state bodies competent in the matters of communications and radio and television broadcasting, the president of the Office of Electronic Communications is a central body of government administration.

58 Pursuant to Article 33 of the Polish Civil Code (Civil Code Act of 23 April 1964 (consolidated text: Journal of Laws of 2020, item 1740, as amended), the State Treasury and organizational units granted a legal personality under specific provisions are bodies corporate.

59 Indirectly, pursuant to Article 8 of the Civil Code, a natural person is a person who has legal capacity from the moment of birth.

providers' obligation to immediately inform the president of the UKE of any change in such details is a logical consequence. This is important because a service provider may not rely on a change of its country representative to *bona fide* third parties, public administration bodies, courts, and prosecutors' offices if the president of the UKE has not been notified and if the change has not been published on the social network site. A change of the country representative in the course of proceedings before public administration bodies, courts, and prosecutors' offices becomes effective upon notification. Information about the country representative is to be published on the social network site in a clearly visible, direct, and permanently accessible manner. On the one hand, such disclosure covers the country representative's full details, including their e-mail address and address for service, and if the country representative is a body corporate, the details of natural persons authorized to act on behalf of that body corporate must also be given. On the other hand, the service provider's full details will also be publicly available, including the full name of the entrepreneur running the social network site or their name and surname, registration or residence address, address for service, registration details, and e-mail address.

A completely different obligation on a service provider contained in the proposed law is to put in place, in the Polish language, effective and understandable internal audit and check processes in matters relating to user complaints concerning, *inter alia*, limited access to content and users' profiles. In this regard, a service provider is obliged to publish the social network site's terms of use in the Polish language and make it available to all users; the terms of use must contain the rules for conducting internal audit and check processes. The terms of use may not be in conflict with the provisions of generally applicable law in Poland. It should therefore not be surprising that in view of the above, a service provider is also obliged to ensure that complaints are sent for the internal audit and check process in a clearly visible and directly and permanently accessible manner. If a user files a complaint, the country representative is required to immediately send confirmation of receipt of the complaint to the e-mail address provided in the complaint. They consider the user's complaint and inform them of the outcome of such consideration via the indicated e-mail address within 48 hours of the filing of the complaint. If the complaint is accepted, the service provider restores the previously limited access to the content or to the user's profile. Information on the outcome of consideration of the complaint should include the reasons for the decision taken, indicating in particular the legal grounds with reference to the social network site's terms of use and the facts serving as the grounds for the decision, along with justification. The information must also contain instructions for exercising the option to file a complaint to the Council, as well as the date and method of the filing of the complaint. On the other hand, if the complaint is not accepted, the service provider informs the user about their option to pursue

their claims through civil action.[60] The president of the UKE supervises the internal audit and check procedure.

60 According to Article 38 of the draft act on freedom of speech: "In the Civil Procedure Code Act of 17 November 1964 (Journal of Laws of 2020, item 1575, 1578 and 2320 and of 2021, item 11), in the first part, in the first book, in title VII, in section VIII, after chapter 1, chapter 2 shall be added to read as follows: 'Chapter 2 Proceedings for the protection of personal rights against persons with unknown identity Article 505^{40}. § 1. The provisions of this chapter shall apply in cases regarding the protection of personal rights, if their violation occurred through the service provider within the meaning of the Electronic Services Act of 18 July 2002 (Journal of Laws of 2020, item 344), and the claimant does not know the first and last name or the name or address of the place of residence or registered office of the respondent who violated their personal rights. § 2. The cases referred to in § 1 are examined by the regional court competent for the place of residence or registered office of the claimant in the territory of the Republic of Poland. § 3. In the cases referred to in § 1, in their claim, instead of the first and last name or the name or address of the place of residence or registered office of the respondent, the claimant provides information enabling the identification of the person who committed the violation, in particular the name of the user profile or login. Article 505^{41} § 1. The claim referred to in Article 505^{40} § 3 shall include: 1) an application for injunction by requiring the service provider referred to in Article 505^{40} § 1 to indicate the respondent's details specified in Article 18(1)–(5) of the Electronic Services Act of 18 July 2002 based on the information provided by the claimant; 2) the identification of the service provider referred to in Article 505^{40} § 1; 3) the indication of publications, in particular written messages, photos, audio and video recordings, which violate the personal rights of the claimant, with the URL of the online data resource on which these were published, the date and time of publication and the name of the user's profile or login, where possible; 4) the claimant's statement that they have attempted to notify the respondent of their intention to bring action against them or a statement that such notification was not possible, along with the reasons. § 2. The claim must be accompanied by a clear reproduction of the publication referred to in § 1(3), in the form of an electronic record and in the form of a printout showing a copied screen image with a visible URL address and the date and time of publication. § 3. The claim referred to in § 1 may contain applications for injunction by: 1) preventing access to the publication, or 2) providing a notice indicating a contentious nature of the publication. Article 505^{42} § 1. The court shall dismiss the claim if: 1) the claim is manifestly unfounded; 2) the action violates the principles of social coexistence or aims to circumvent the law, in particular the provisions of the Personal Data Protection Act of 10 May 2018 (Journal of Laws of 2019, item 1781). § 2. The court shall dismiss the claim in closed session. § 3. The justification of the decision shall be made *ex officio* in writing. It shall only contain an explanation as to why the claim was considered manifestly unfounded or violates the principles of social coexistence or aims to circumvent the law, in particular the provisions of the Personal Data Protection Act of 10 May 2018. The decision, along with the justification, shall be delivered by the court *ex officio* only to the claimant, with an instruction on the manner and time limit for filing an appeal. Article 505^{43} § 1. The court shall request the service provider to send in the data referred to in Article 505^{41} § 1(1). § 2. The court shall discontinue the proceedings if the service provider fails to send in the data referred to in Article 505^{41} § 1(1) within 3 months of the delivery of the court's request for data. The decision shall be served only on the claimant. The decision shall be accompanied by an instruction on the option to bring action under generally applicable laws and regulations. § 3. The claimant may file a complaint to any such decision. § 4. The complaint shall be examined by the court which issued the challenged decision, sitting in a bench of three judges. § 5. If the service provider has failed to send in the data referred to in Article 505^{41} § 1(1) without justified reasons, the court shall sentence them to a fine according to the provisions on penalties for a witness' failure to attend. Where the service provider is a body corporate or an organizational unit, a fine shall be imposed on the manager or employee responsible for fulfilling the obligation to provide the data. Article 505^{44}. The court shall examine the case according to the general provisions after the service provider has sent in the data referred to in Article 505^{41} § 1(1).'"

The draft act on freedom of speech also sets out the procedure before the Council. A user who is dissatisfied with the decision made regarding their complaint during the internal audit and check process may submit a complaint to the Council within seven days of receiving information about the outcome of the consideration of their complaint. The user files a complaint with the Council, along with information about the decision on the complaint filed with the service provider and confirmation of receipt thereof in the form of an electronic document uploaded to the public administration services' electronic platform.[61] The complaint is to be signed by the party or its statutory representative or proxy with a qualified electronic signature, a trusted signature, or a personal signature. Upon the ineffective expiry of the time limit for consideration of the complaint under the internal audit and check process, the user may file a complaint with the Council without attaching information about the decision on the complaint filed with the service provider. The complaint filed with the Council must include the first and last name or the name of the complainant and their legal representative or proxy, if appointed; identification of the service provider and its country representative; a description of the violation; and indication of the request. The parties to the proceedings before the Council are the user whose access to content was limited or the user whose access to their profile was limited and the service provider. Importantly, during the proceedings before the Council, the service provider acts through its country representative. Where a complaint has been filed, the Council will immediately notify the country representative of the user's complaint. The country representative provides the Council with the materials collected during the internal audit and check process within 24 hours of notification of the complaint filed by the user. What is crucial, as a result of the conducted proceedings, is that the Council issues a decision in which it either orders the restoration of the previously limited access to content or to the user's profile, if it finds that the content or user profile to which access was limited does not constitute illegal content,[62] or it refuses to restore the previously limited access to the content or to the user's profile, if it finds that the content or user's profile to which access was limited constitutes illegal content. It should be noted that the above-indicated feature of illegality refers to a violation of Polish law. The Council issues a decision within seven days of the date of receipt of the complaint.

The evidence taken before the Council is limited to the evidence the user provided with the complaint, evidence provided by the country representative, and evidence that can be ascertained on the basis of the data available to the Council

61 Pursuant to Article 3 of the Act of 17 February 2005 on the computerization of the activities of entities performing public tasks (consolidated text: Journal of Laws of 2021, item 670), an electronic platform of public administration services is an ICT system in which public institutions provide services through a single access point on the Internet.

62 As a reminder, according to Article 3(8) of the draft act on freedom of speech, the term 'illegal content' is understood as content that violates personal rights, disinformation, or content of criminal nature, as well as content that violates morality, in particular, by disseminating or praising violence, distress, or humiliation.

itself. The Council does not take evidence from the testimony of witnesses, hearing of the parties, expert opinions, or site inspections. It should be noted that of great importance here will be the use of electronic evidence, which has traditionally been used in civil proceedings, for example, to support the material circumstances of infringement on personal rights on the Internet,[63] and in criminal proceedings, for example, to combat cybercrime.[64] In this case, however, its potential will be used to prove the occurrence of unlawful censorship on the Internet affecting Polish users. In view of the above, a user is obliged to present all the evidence at the latest upon filing the complaint, under pain of losing the right to present such evidence in the course of further proceedings. Evidence submitted in breach of this obligation is disregarded, unless the party substantiates that its submission in the complaint was not possible. The Council disregards any evidence that would lead to an extension of the proceedings. On the other hand, the justification of the decision issued in the proceedings before the Council may be limited to indicating the facts which the Council considers proven and quoting the legal provisions constituting the legal basis for the decision. The Council delivers the decision to the parties immediately, no later than within 24 hours of issuance, and the service provider is required to execute the Council's decision immediately, no later than within 24 hours of delivery. Within another 24 hours from the expiry of the time limit for the execution of the decision, the country representative informs the Council about the manner of performance. If the service provider fails to comply with the decision, the Council immediately notifies the president of the UKE. Importantly, the service provider cannot once again limit access to content that was the subject of examination before the Council.

The draft act on freedom of speech also provides for numerous administrative penalties in the event of failure by a service provider or its country representative to comply with the obligations set out in this proposed legislation. For example, a service provider who, *inter alia*, fails to fulfil the obligation to submit the report; appoint the country representative; immediately notify the president of the UKE of the appointment or change of its country representative or provide their details; put in place an effective and understandable internal audit and check process in the Polish language; post the social network site's terms of use, including the internal audit and check process, on the site in the Polish language so that it is available to all users; provide a clearly visible and directly and permanently available mechanism for submitting complaints through the internal audit and check process; or implement any

63 Judgement of the Court of Justice of the European Union of 17 October 2017 in case C-194/16…; Judgement of the Court of Justice of the European Union of 25 October 2011 in joined cases C-509/09 and C-161/10…

64 For example, see: Gercke, 2004, p. 802; Winick, 1994, pp. 75–128; Katyal, 2001, pp. 1003–1114; Marion, 1997, pp. 67–108; Silver, 2001, p. 5; Sinrod and Reilly, 2000, pp. 1–53; Speer, 2000, pp. 259–273; Gercke, 2009, pp. 409-420; Brenner and Schwerha, 2004, pp. 111–114; Hilley, 2005, pp. 171–174; Walden, 2004, pp. 321–336; Moitra, 2005, pp. 435–464; Wang, 2007, pp. 216–223; Nuth, 2008, pp. 437–446; Chung et al., 2006, pp. 669–682; Boni, 2001, pp. 18–19; Clough, 2014, pp. 698–736.

Council decision to restore previously limited access to content or to a user's profile is liable to a financial penalty. Further, a service provider is liable to a financial penalty for its country representative's failure to fulfil the obligation to consider a user complaint filed under the internal audit and check process, provide responses and any information to institutions and authorities in relation to any conducted proceedings, or participate in training organized by the president of the UKE on the current legal status regarding complaints considered through the internal audit and check process. For the service provider's breach of each of these obligations, the Council imposes, by way of a decision, a financial penalty in the amount of PLN 50,000 to PLN 50,000,000. When imposing a financial penalty, the Council considers the impact of the service provider's omission on the magnitude of disinformation caused, the degree of breach of the public interest, the frequency of past breaches of an obligation or breaches of a prohibition of the same type as the breach of the present obligation or breach of prohibition as a result of which the penalty is to be imposed, and the past record of penalties for the same conduct and actions voluntarily taken by the party to avoid the consequences of a violation of the law. A party dissatisfied with a Council decision may apply to the Council for reconsideration of the case, so it does not have a right to appeal.[65]

The above is a presentation of the scope of the regulations contained in the draft act on freedom of speech with regard to the issue signaled in the title, i.e., issues of unlawful censorship. Nevertheless, it should be clearly noted that this draft legal act also has another dimension, which is to prevent and combat false information posted on the Internet, for example, for the purpose of disinformation or discrediting. This is an important issue that will be discussed in another chapter in this monograph. With regard to the issue of censorship, it is clear that the main objective of the draft act on freedom of speech is to counteract and combat unlawful censorship. The above-presented regulations apply to it. It is the main problem that the proposed legislation is supposed to solve. This is because lawful censorship, as one sanctioned in generally applicable law in Poland, does not fall within the scope of the application of the draft act on freedom of speech. Therefore, the example of censorship under copyright laws in Poland has a completely different legal status.

Certainly, the fact that a draft act on freedom of speech has been proposed for adoption in Poland as universally binding law should be assessed in very positive terms. Lawyers, both scholars and practitioners, cannot deny the threats resulting from the current technical, technological, and civilizational progress. It is a matter of a simple causal relationship. Since a new level of human activity has emerged in

65 As a rule, pursuant to Article 127 § 1 of the Code of Administrative Procedure (Code of Administrative Procedure Act of 14 June 1960 (consolidated text: Journal of Laws of 2021, item 735)), a party may appeal against a decision issued in the first instance on one occasion only. On the other hand, pursuant to §3 of the same provision, a decision issued in the first instance by a minister or a local government court of appeals is not subject to appeal; however, a party who is dissatisfied with the decision may apply to that body for reconsideration of the case, and the provisions on appeals against decisions will apply *mutatis mutandis* to that application.

the form of cyberspace, we must be aware that cybercrime is concomitant. Since cybercrime is emerging, we should introduce legal regulations regarding electronic evidence and technologically advanced evidence taking. This is exactly the case with the subject indicated in the title. Since the Internet appears to have become a new and increasingly popular space for human activity, we must be aware that numerous threats to individual rights and freedoms will also arise. Online censorship is one such threat. Since there is a phenomenon of censorship on the Internet, we must be aware of it and be in a position to distinguish between lawful and unlawful censorship. Since we can see and understand unlawful censorship on the Internet, we can counteract it by introducing a law such as the draft act on freedom of speech in Poland.

The quality of the proposed law deserves high praise. The regulation is comprehensible, transparent and, in principle, unconditional. Thus, alongside professional lawyers, average users will also understand the proposed principles of freedom of speech on the Internet. This is what law should be about in the first place. Second, it should be noted that although there has not yet been an opportunity to test this legislation in practice, it seems to be adequate and justified. It is adequate, meaning it does not go beyond what is necessary, and it protects what is necessary. For example, the administrative penalty range is from PLN 50,000 to PLN 50,000,000, which is wide enough to correspond to the variety of digital platforms and social media on the Internet. Furthermore, the legislation protects individuals' freedoms and rights, which in themselves entail a need for protection. On the other hand, it is justified in that it responds to the actual phenomenon of unlawful censorship, which must be counteracted in line with the essence of the rights and freedoms of every human being, as universally recognized in all civilized countries in today's world. Equally important is the fact that the legal norms contained in the draft act on freedom of speech, by filling the regulatory gap, have systemic consistency with other legal acts in force in Poland. There seems to be no contradiction or logical inconsistency between them, especially due to the fact that the proposed law adapts the wording of other legal acts that are generally applicable in Poland.[66] Consequently, the assessment of systemic compatibility is satisfactory. Ultimately, the axiology of law is also important. The draft act on freedom of speech is an equitable law, based on the principle of justice, including social justice, and it strives to implement the value of truth,[67] duly taking account of the dignity of every human being, which is the source of their rights and freedoms, and introducing justified controls on the Internet.

66 For example, according to Article 39 of the draft act on freedom of speech, "in the Act of 31 July 1981 on the remuneration of persons holding state managerial positions (Journal of Laws of 2020, item 1637) in Article 2(4), after the words 'a member of the State Commission for the investigation of cases of acts against sexual freedom and morals towards a minor under the age of 15,' the words 'a member of the Freedom of Speech Council' shall be added."

67 See Wielec, 2017, pp. 149–277; Judycki, 2001, pp. 25–26; Zajadło, 2013, pp. 20–32; Waltoś, 2014, pp. 273–281; Jodłowski, 2015, pp. 54–71; Dębowski, 2014, pp. 12–15; Strogowicz, 1959, p. 85; Murzynowski, 1976, p. 131; Jabłońska-Bonca, 1999, p. 80; Klejnowska, Kłak, and Sobolewski, 2011, p. 45; Dyl A. et al., 2012, pp. 23–24.

To summarize, the draft act on freedom of speech proposed by the Polish Ministry of Justice consists of a number of provisions protecting freedom of speech against unlawful censorship on the Internet. It defines the rules for controlling activities involving the provision of electronic services via social network sites. It defines service providers' obligations as regards the guarantees of the right to freedom of expression. It lays down rules for internal audit and check processes to be performed by service providers with regard to user complaints concerning violations of the right to freedom of expression. It contains provisions on proceedings before public administration authorities and court actions in the event of restriction of access to an electronic service provided via an online social network site. It also provides for sanctions. Taking all the reasons cited above into account, the law's motives are highly rated. This law is Poland's response to unlawful censorship on the Internet and may serve as a model for imitation or inspiration for other countries, not only in Europe, but in general.

4. Freedom of speech and pluralism in Poland

In Poland, freedom of speech and pluralism are respected as individuals' constitutionally guaranteed rights and freedoms. Pursuant to Article 14 of the Polish Constitution, Poland ensures freedom of the press and other means of social communication. Pursuant to Article 49 of the Polish Constitution, the freedom and privacy of communication are ensured. Any limitation thereon may be imposed only in cases and in a manner specified by statute. Pursuant to Article 53 of the Polish Constitution, freedom of conscience and religion shall be ensured to everyone. Freedom of religion includes the freedom to profess or to accept a religion by personal choice as well as to manifest such religion, either individually or collectively, publicly or privately, by worshipping, praying, participating in ceremonies, and performing rites or teaching. Freedom of religion also includes possession of sanctuaries and other places of worship for the satisfaction of believers' needs, as well as the right of individuals, wherever they may be, to benefit from religious services. In this context, parents have the right to ensure their children a moral and religious upbringing and teaching in accordance with their convictions.[68] The religion of a church or other legally recognized religious organization may be taught in schools, but other peoples' freedom of religion and conscience may not be infringed thereby. The freedom to publicly express religion may be limited only by means of statute and only where

68 In this case, the provision of Article 48(1) of the Polish Constitution applies *mutatis mutandis*. According to it, parents have the right to rear their children in accordance with their own convictions. Such upbringing should respect the child's degree of maturity as well as their freedom of conscience and belief and also their convictions.

this is necessary for the defense of state security, public order, health, morals, or the freedoms and rights of others. No one may be compelled to participate or refrain from participating in religious practices. No one may be compelled by organs of public authority to disclose their philosophy of life, religious convictions, or belief. Pursuant to Article 54 of the Polish Constitution, the freedom to express opinions and acquire and disseminate information is ensured to everyone. Preventive censorship of the means of social communication and the licensing of the press are prohibited. Statutes may require the receipt of a permit for the operation of a radio or television station. Pursuant to Article 73 of the Polish Constitution, the freedom of artistic creation and scientific research, as well as dissemination of the fruits thereof, and the freedom to teach and enjoy the products of culture are ensured to everyone.

In the opinion of the Polish Constitutional Court (CC), freedom of speech covers all levels of an individual's activity, is an expression of the dignity of a person's autonomy, and creates opportunities for full personality development in the individual's cultural and civilizational environment.[69] However, the most important thing in the context of the subject matter covered in the title is the correct understanding of the normative content of Article 54 of the Polish Constitution. One of the CC's rulings[70] underlined that Article 54(2) of the Polish Constitution states that "preventive censorship of the means of social communication and the licensing of the press are prohibited." It noted that the provision contains an absolute norm related to the essence of freedom of all expression in these domains. The Court concluded that the prohibition of the licensing of the press is closely related to the need to make freedom of speech a reality. It pointed out that it is rightly emphasized in the doctrine that the effective pursuit of the freedom of the press must be based on the principle of freedom and the liberal model, while any prevention based on censorship or prior authorization is incompatible with these ideas.[71] In this regard, it stressed that censorship or a preventive system of press control interferes with the essence of the constitutional freedom of expression by means of social communication. The key point, however, is that in its ruling, the CC interpreted the normative content of Article 54 of the Polish Constitution. By comparing the meaning of Article 54 of the Polish Constitution and Article 10 of the Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR),[72] it indicated that the former's meaning is narrower than that provided for in the latter. According to the CC, this is so because Article 10 of the ECHR concerns the freedom of all expression, which covers not only the expression of views through speech, but also in written, printed, or artistic form. Further, the presented ruling stressed that Article 54 of the Polish Constitution provides for the freedom to express one's views and to source and

69 Judgement of the Constitutional Court of 12 May 2008, case ref. SK 43/05; Judgement of the Constitutional Court of 14 December 2011, case ref. SK 42/09.

70 Judgement of the Constitutional Court of 20 February 2007, case ref. P 1/06.

71 Cited from the CC: J. Israel, *Droit des libertés fondamentales* (Paris: éd. L.G.D.J.,1988), p. 463.

72 Convention for the Protection of Human Rights and Fundamental Freedoms (4 November 1950, as amended), (Polish Journal of Laws of 1993, No. 61, item 284).

disseminate information in verbal, written, and printed form, including through any means of social communication. Moving on to the key conclusion, the CC stated that the acquisition and dissemination of information referred to in Article 54 of the Constitution comprise not only the freedom to source and disseminate facts, but also the views and opinions of others. Therefore, in its opinion, the normative content of Article 54 of the Polish Constitution is contained in the democratic standard of freedom of expression as defined today, and in particular, in the generally defined Article 10 of the ECHR.

In its entirety, the above argument clearly demonstrates that freedom of expression is provided for in the Polish Constitution, while the same legal norms contain the principle of pluralism in this respect. However, this was not achieved in a single provision, but in many. Only a comparison of Articles 14, 49, 53, 54, and 73 of the Polish Constitution, with their proper interpretation, in particular of Article 54, confirms this conclusion. Therefore, it should not be surprising that, in accordance with Polish law, it is only admissible to limit these individual rights and freedoms in such a way that does not infringe on their essence. The normative basis for any such limitation is provided for in Article 31(3) of the Polish Constitution. Pursuant to this provision, any limitation upon the exercise of constitutional freedoms and rights may be imposed only by statute and only when necessary in a democratic state for the protection of its security or public order or to protect the natural environment, health, or public morals, or the freedoms and rights of other persons. These limitations may not infringe on the substance of the freedoms and rights. The scope of Article 31(3) of the Polish Constitution is universal, as it applies to all freedoms and rights guaranteed in the Constitution.[73] In the CC's opinion, the principle of proportionality dictates: the application of measures that enable the effective achievement of the intended objectives; the application of only necessary measures, i.e., measures that are the least burdensome for the individual; and the maintenance of an appropriate balance between the benefit of the measures applied and the burden imposed on the individual.[74] Thus, any provision introducing a limitation is inconsistent with the Polish Constitution if the same effects could be achieved by means that restrict the exercise of personal freedoms or rights to a lesser extent.[75] This means that in Poland, the limitations under Articles 14, 49, 53, 54, and 73 of the Polish Constitution may be imposed, but solely on strictly defined principles resulting from Article 31(3) of the Polish Constitution. These principles fall into two dimensions. The formal dimension requires that the limitations be introduced in the form of a statute (act), and the material dimension requires that they be necessary in a democratic state for the protection of its security or public order, or to protect the natural environment,

73 Judgement of the Constitutional Court of 26 May 2008, case ref. SK 25/07.

74 Judgement of the Constitutional Court of 28 June 2000, case ref. K 34/99; Judgement of the Constitutional Court of 29 September 2008, case ref. SK 52/05.

75 Judgement of the Constitutional Court of 30 October 2006, case ref. P 10/06.

health, or public morals, or the freedoms and rights of other persons. Both of these dimensions must be met jointly.

5. Conclusions

It should be noted that law undergoes substantial changes. It is clear that not only the legislation is subject to change, but also its interpretation. It seems that it is equally clear that reality can be ahead of law and thus force the introduction of new legislation. This is precisely the case, for both aspects, with the subject indicated in the title. First, as a result of CJEU case law, the construction of Article 3(1) of Directive 2001/29/EC has been altered. Second, the fact of controlling and limiting Internet users' activities, including their publication of content, in particular on political or moral grounds but not legal grounds, forces the wise legislator to take appropriate, targeted, and effective legislative action. In order to determine whether the analyzed cases and forms of censorship represent limitations under the principles provided for in Article 31(3) of the Polish Constitution, it is necessary to determine whether they meet the formal and material aspects of this legal norm.

Lawful censorship constitutes a direct restriction of the rights and freedoms provided for in the Polish Constitution. The presented example of such censorship results from the implementation of Article 3(1) of Directive 2001/29/EC in the Polish normative system. This transposition of EU law was made in the form of a statute (act). This means that the formal aspect of Article 31(3) of the Polish Constitution was followed correctly. Incidentally and to signal an issue, an interesting thing here is the problem of the CJEU's subsequent modification of the normative meaning of Article 3(1) of Directive 2001/29/EC in the context of the requirement to keep the statutory form for restrictions on individual rights and freedoms in Poland. At this point, however, one cannot but assume that the CJEU decoded the proper meaning of the EU concept of 'communication to the public' without creating new standards, to only present those that are not visible at first sight in the content of the provision, being hidden in the essence of the legal norm. In this situation, it is crucial to determine whether Article 3(1) of Directive 2001/29/EC, as implemented in the Polish legal system and then enriched with the CJEU's autonomous constructions, meets the requirements of the material aspects of Article 31(3) of the Polish Constitution, that is, whether such a system of the admissibility of posting content on the Internet through the perspective of copyright laws is necessary in a democratic state for the protection of its security or public order, or to protect the natural environment, health, or public morals, or the freedoms and rights of other persons. Obviously, copyright protection is about others' freedoms and rights, that is, the copyright holder's legitimate claims. Nevertheless, as indicated above, these restrictions may not infringe on the essence of other freedoms and rights. Therefore, for them to be

considered reasonable, justified, and lawful, one must first determine whether the restrictions enable the effective achievement of the objectives pursued. It seems that the system for the admissibility of hyperlinking, as defined by the CJEU, ensures full achievement of the objectives set, i.e., copyright protection on the Internet. Second and third, one should examine whether such stringently defined hyperlinking rules constitute a necessary measure, i.e., the least burdensome to an individual out of the available measures, and whether an appropriate proportion was maintained between the benefit of the measures applied and the burden imposed on an individual. Over these points, the answer is not so obvious and requires an in-depth and multi-contextual analysis, that is, an analysis that will not focus on generalities, but will examine each CJEU ruling separately. However, this requires a much broader scholarly presentation. At present, what can be stated with certainty is that the conditions of compatibility assessment will follow different paths for the judgements in the GS Media case and related rulings, and for the judgements in the Renckhoff case and related rulings. These are clearly separate groups of judgements.

In the case of unlawful censorship, it is obvious that it openly violates and restricts the individual rights and freedoms provided in the Polish Constitution. This is one of two indisputable facts. The other one is that it has no basis in generally applicable law in Poland. This means that, from the outset, this type of censorship does not meet the requirements of the formal aspect of Article 31(3) of the Polish Constitution. The conclusion is therefore that unlawful censorship constitutes a limitation—without a legal sanction—of the individual rights and freedoms provided in the Polish Constitution. In this context, the Polish draft act on freedom of speech should be treated as a wise response to this negative phenomenon that has emerged and is growing on the Internet.

The protection of freedom of expression and the preservation of pluralism on the Internet constitute a new objective in law for the contemporary legislator. It is an opportunity to preserve the standards of human rights protection developed over the years in the real world in the realm of the digital world. After all, to opt for progress does not mean forgetting the human values, ideals, or principles universally recognized by civilized nations. Otherwise, the progress is spurious.

Bibliography

Afsarmanesh, H., Masís, V., Hertzberger, L. (2004) 'Virtual Community Support in Telecare' in Camarinha-Matos, L., Afsarmanesh, H. (eds.) *Processes and Foundations for Virtual Organizations: IFIP TC5/WG5.5 4th Working Conference on Virtual Enterprises, PRO-VE 2003*. Boston, Springer, pp. 211–220.

André, A. (2019) 'The Information Technology Revolution in Health Care' in André A. (ed.) *Digital Medicine. Health Informatics*. Cham: Springer, p. 4.

Boni, B. 'Creating a Global Consensus Against Cyber Crime', *Network Security*, 2001/9, pp. 18–19.

Brenner, S., Schwerha, J. (2004) 'Introduction—Cybercrime: A Note on International Issues', *Information Systems Frontiers*, 6(2), pp. 111–114.

Chung, W., Chen, H., Chang, W., Chou, S. 'Fighting Cyber Crime: A Review and the Taiwan Experience', *Decision Support Systems*, 2006/41, pp. 669–682.

Clough, J. (2014) 'A World of Difference: The Budapest Convention on Cybercrime and the Challenges of Harmonization', *Monash University Law Review*, 40(3), pp. 698–736.

Colangelo, G., Maggiolino, M. 'ISPs' Copyright Liability in the EU Digital Single Market Strategy', *International Journal of Law and Information Technology*, 2018/3, pp. 142–159.

de Pietro, C., Francetic, I. (2018) 'E-health in Switzerland: The Laborious Adoption of the Federal Law on Electronic Health Records (EHR) and Health Information Exchange (HIE) Networks', *Health Policy*, 122(2), pp. 69–74.

Dębowski, J. (2014) 'O klasycznej koncepcji prawdy i jej filozoficznych podstawach. Czy w Matrixie możliwa jest prawda?' in Kilkiewicz, A., Starzyńska-Kościuszko, E. (ed.) *Oblicza prawdy w filozofii, kulturze, języku*. Olsztyn: Instytut Filozofii Uniwersytetu Warmińsko-Mazurskiego, pp. 12–15.

Doroszewski, W. (ed.) *Słownik Języka Polskiego* [Online]. Available at: https://sjp.pwn.pl/doroszewski/cenzura;5416093.html (Accessed: 15 May 2021).

Držanič, I., Mladenović, V., Debevc, M., Dolničar, V., Petrovčič, A., Touzery, S., Kožuh, I. (2019) 'Usability Testing of a Smartphone Telecare Application for Informal Caregivers' in Milošević, D., Tang, Y., Zu Q. (eds.) *Human Centered Computing: 5th International Conference, HCC 2019*. Cham, Springer, p. 252.

Dutton, W. 'Internet Studies: Foundations of a Transformative Field' in Dutton, W. (ed.) *The Oxford Handbook of Internet Studies*. Oxford: Oxford University Press, 2013, p. 1.

Dyl, A., Dyl, K., Miętkowska, K., Tokarz, A. *Procedura karna. Repetytorium*. Warszawa: Wolters Kluwer, 2012, 23–24.

Elenko, E., Underwood, L., Zohar, D. (2015) 'Defining Digital Medicine', *Nature Biotechnology*, 33(5), pp. 456–461.

Eurostat (25 May 2021) 'Individuals – internet use'. Available at: https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ifp_iu/default/table?lang=en (Accessed: 20 April 2021).

Facebook Terms of Service [Online]. Available at: https://www.facebook.com/legal/terms (Accessed: 20 May 2021).

Fernández-Díez, I. 'La puesta a disposición de obras y prestaciones protegidas por quien no tiene la autorización de los titulares de derechos de propiedad intelectual cuando previamente se ha hecho accesible dicha obra en Internet con licencia del autor', *La Ley Unión Europea*, 2018/64, p. 2.

GAO, X., Lin, L., Lan, T., Gan, X. (2020) 'Design and Research on the Chinese Medicine Health Management System Based on the Wireless Sensor Network' in Xu, Z., Choo, K., Dehghantanha, A., Parizi, R., Hammoudeh, M. (eds.) *Cyber Security Intelligence and Analytics*. Advances in Intelligent Systems and Computing 928. Cham, Springer, pp. 55–56.

GERCKE, M. 'Europe's Legal Approaches to Cybercrime', *ERA Forum*, 2009/10, pp. 409–420.

GERCKE, M. 'Analyse des Umsetzungsbedarfs der Cybercrime Konvention. Teil 2: die Umsetzung im Bereich des Strafverfahrensrechts', *Multimedia und Recht*, 2004/12, pp. 801–806.

GINSBURG, J. (2017) 'The Court of Justice of the European Union Creates an EU Law of Liability for Facilitation of Copyright Infringement: Observations on Brein v. Filmspeler [C-527/15] (2017) and Brein v. Ziggo [C-610/15] (2017)' Columbia Law and Economics Working Paper 572.

HELIOS, J. Jedlecka W. (2018) *Wykładnia prawa Unii Europejskiej ze stanowiska teorii prawa*. Wrocław: Wydział Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego.

HILLEY, S. 'Pressure Mounts on US Senate to Pass Cybercrime Treaty', *Digital Investigation*, 2005/2, pp. 171–174.

HUANG, C. (2009) 'Medical Informatics' in Nof, S. (ed.) *Springer Handbook of Automation*. Berlin: Springer, 1423–1432.

ISTEPANIAN, R., LAXMINARAYAN, S., PATTICHIS, C. (eds.) (2006) Preface in *M-Health. Emerging Mobile Health System.* Cham, Springer.

JABŁOŃSKA-BONCA, J. (1999) *O prawie, prawdzie i przekonywaniu.* Koszalin: Wydawnictwo Uczelniane Bałtyckiej Wyższej Szkoły Humanistycznej.

JODŁOWSKI, J. *Zasada prawdy materialnej w postępowaniu karnym. Analiza w perspektywie funkcji prawa karnego.* Warszawa, 2015, pp. 54–71.

JUDYCKI, S. (2001) 'O klasycznym pojęciu prawdy', *Roczniki Filozoficzne*, 49(1), pp. 25–62.

KATYAL, N. (2001) 'Criminal Law in Cyberspace', *University of Pennsylvania Law Review*, 149(4), pp. 1003–1114.

KLAR, R., PELIKAN, E. (2011) 'Telemedicine in Germany' in Kramme, R., Hoffmann, K., Pozos R. (eds.) *Springer Handbook of Medical Technology.* Berlin: Springer, 1119–1127.

KLEJNOWSKA, M., KŁAK, C., SOBOLEWSKI, Z. (2011) *Proces karny. Część ogólna*. Warszawa: Wolters Kluwer.

KOO, J. (2018) 'Away We Ziggo: The Latest Chapter in the EU Communication to the Public Story', *Journal of Intellectual Property Law & Practice*, 13(7), pp. 542–551.

LINKOUS, J. (2001) 'A Rapidly Evolving Definition of Telemedicine' in Kiel J. (ed.) *Information Technology for the Practicing Physician. Health Informatics*. New York: Springer, p. 226.

LONG, J. (2018) 'Different Solutions for Similar Questions: Hyperlinks and the Right of Communication to the Public in China and the EU', *International Review of Intellectual Property and Competition Law*, 49, pp. 413–435.

LUPTON, D. (2013) 'The Digitally Engaged Patient: Self-Monitoring and Self-Care in the Digital Health Era', *Social Theory & Health*, 11(3), pp. 256–270.

MAIMONE, A., BIDWELL, J., PENG, K., FUCHS, H. (2012) 'Augmented Reality: Enhanced Personal Autostereoscopic Telepresence System Using Commodity Depth Cameras', *Computers & Graphics*, 36(7), pp. 791–807.

MARION, N. (1997) 'Symbolic Policies in Clinton's Crime Control Agenda', *Buffalo Criminal Law Review*, 1(1), pp. 67–108.

MARS, M., SCOTT, R. (2010) 'Global E-health Policy: A Work in Progress', *Health Affairs*, 29(2), pp. 237–243.

MOITRA, S. (2005) 'Developing Policies for Cybercrime', *European Journal of Crime, Criminal Law and Criminal Justice*, 13(3), pp. 435–464.

MURZYNOWSKI, A. *Istota i zasady procesu karnego.* Warszawa: PWN, 1976.

NORDEMANN, J. (2018) 'Recent CJEU Case Law on Communication to the Public and Its Application in Germany: A New EU Concept of Liability', *Journal of Intellectual Property Law & Practice*, 13(9), pp. 744–756.

NUTH, M. (2008) 'Taking Advantage of New Technologies: For and Against Crime Computer Law and Security Report', *Computer Law & Security Review*, 24(5), pp. 437–446.

ORĘZIAK, B. (2017) 'Judgment of the Court of Justice of the European Union (Second Chamber) of 8 September 2016 in Case C 160/15 in the proceedings of GS Media v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruid Dekker', *Comparative Law Review*, 23(1), pp. 243–253.

ORĘZIAK, B. (2018a) 'Czy linkowanie w sieci jest dopuszczalne? Analiza w świetle najnowszego orzecznictwa Trybunału Sprawiedliwości Unii Europejskiej' in Żak D. (ed.) *Własność intelektualna w Internecie.* Lublin: Towarzystwo Naukowe KUL, Katolicki Uniwersytet Lubelski, pp. 137–157.

ORĘZIAK, B. 'Ograniczenia linkowania w Internecie w świetle orzecznictwa Trybunału Sprawiedliwości Unii Europejskiej – analiza orzeczenia z dnia 14 czerwca 2017 roku w sprawie C-610/15', *Przegląd Prawno-Ekonomiczny*, 2018b/4, pp. 199–219.

ORĘZIAK, B. 'Analiza prawnej dopuszczalności zamieszczania hiperlinków w internecie. Uwagi na tle wyroku Europejskiego Trybunału Praw Człowieka z 4.12. 2018 r. w sprawie MAGYAR JETI ZRT przeciwko Węgrom', *Prawo w Działaniu Sprawy Cywilne*, 2019a/40, pp. 181–192.

ORĘZIAK, B. (2019b) 'Judicial Dialogue between the European Court of Human Rights and the Court of Justice of the European Union in the Field of Legal Liability for Posting Hyperlinks', *International Community Law Review*, 21(5), pp. 432–448.

PAGLIALONGA, A., MASTROPIETRO, A., SCALCO, E., RIZZO, G. (2019) 'The mHealth' in Andreoni, G., Perego, P., Frumento, E. (eds.) *m_Health Current and Future Applications*. Cham: Springer, pp. 5–17.

RADOSAVLJEV, P. (2017) 'For Whom the Copyright Scale Tips: Has the CJEU Established a Proper Balance of Rights with Its GS Media Case Judgement?', *European Journal of Law and Technology*, 8(3).

RADZIEWICZ, P. 'Wzruszenie "domniemania konstytucyjności" aktu normatywnego przez Trybunał Konstytucyjny', *Przegląd Sejmowy*, 2008/5, pp. 55–86.

ROSATI, E. (2017) 'GS Media and Its Implications for the Construction of the Right of Communication to the Public within EU Copyright Architecture', *Common Market Law Review*, 54(4), pp. 1221–1242.

ROSENMEIER, M., SZKALEJ, K., WOLK, S. (2019) *EU Copyright Law: Subsistence, Exploitation and Protection of Rights.* Alphen aan den Rijn: Kluwer.

SEZGIN, E. (2018) 'Introduction to Current and Emerging mHealth Technologies: Adoption, Implementation, and Use' in Sezgin, E., Yildirim, S., Yildirim, S., Özkan-Yildirim, S., Sumuer, E. (eds.) *Current and Emerging mHealth. Technologies Adoption, Implementation, and Use*. Cham: Springer, pp. 1–6.

SILVER, O. 'European Cyber Crime Proposal Released', *Computer Fraud and Security*, 2001/5, pp. 3–17.

SINROD, J., REILLY, W., (2000) 'Cyber-Crimes: A Practical Approach to the Application of Federal Computer Crime Laws', *Santa Clara Computer and High Technology Law Journal*, 16(2): 1–53.

Słownik Języka Polskiego [Online]. Available at: https://sjp.pwn.pl/sjp/cenzura;2447537.html (Accessed 15 May 2021).

SPEER, L. 'Redefining Borders: The Challenges of Cybercrime', *Crime, Law and Social Change*, 2000/34, pp. 259–273.

STROGOWICZ, M. (1959) *Prawda obiektywna i dowody sądowe w radzieckim procesie karnym*. Warszawa: Wydawnictwo Prawnicze.

TERRY, N. (2000) 'Structural and Legal Implications of E-health', *Journal of Health Law*, 33(4), pp. 605–614.

VISSER, D. 'Hof van Justitie van de Europese Unie, 7 augustus 2018, C-161/17, ECLI: *EU: C: 2018:* 634', *AMI: tijdschrift voor auteurs-, media-en informatierecht*, 2018/8, 183–190.

VISSER, D. 'Kroniek van de intellectuele eigendom', *Nederlands Juristenblad*, 2018/93, pp. 1025–1037.

WALDEN, I. (2004) 'Harmonising Computer Crime Laws in Europe', *European Journal of Crime, Criminal Law and Criminal Justice*, 12(4), pp. 321–336.

WALTOŚ, S. (2014) 'Zasada prawdy materialnej' in Wiliński P. (ed.) *System Prawa Karnego Procesowego*. Vol. 3, *Zasady procesu Karnego*. Warszawa: Wolters Kluwer, 273–281.

WANG, F. 'Current Developments – Europe', *Intellectual Property Forum: Journal of the Intellectual and Industrial Property Society of Australia and New Zealand*, 2018/113, pp. 86–91.

WANG, S. (2007) 'Measures of Retaining Digital Evidence to Prosecute Computer-based Cyber-crimes', *Computer Standards and Interfaces*, 29(2), pp. 216–223.

WANG, X., LOVE, P., KIM, M., WANG, W. (2014) 'Mutual Awareness in Collaborative Design: An Augmented Reality Integrated Telepresence System', *Computers in Industry*, 65(2), pp. 314–324.

WIELEC, M. (2017) *Wartości – analiza z perspektywy osobliwości postępowania karnego*. Lublin: Academicon.

WINICK, R. (1994) 'Searches and Seizures of Computers and Computer Data', *Harvard Journal of Law & Technology*, 8(1), pp. 75–128.

ZAJADŁO, J. 'Teoretyczne i filozoficzno-prawne pojęcie prawdy' in Kremens, K., Skorupka J. (eds.) *Pojęcie, miejsce i znaczenie prawdy w procesie karnym*. Wrocław: Oficyna Prawnicza, 2013), 20–32.

ZAWIDZKA-ŁOJEK, A. Grzeszczak, R. (2015) *Prawo materialne Unii Europejskiej*. Warszawa: Instytut Wydawniczy EuroPrawo.

Legal sources

Decree of 5 July 1946 establishing the Main Office for the Control of the Press, Publications, and Public Performances (Journal of Laws of 1946, No. 34, item 210).

Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, as amended (Journal of Laws of 1993, No. 61, item 284).

Code of Administrative Procedure Act of 14 June 1960 (consolidated text: Journal of Laws of 2021, item 735).

Civil Code Act of 23 April 1964 (consolidated text: Journal of Laws of 2020, item 1740, as amended).

Code of Civil Procedure Act of 17 November 1964 (consolidated text: Journal of Laws of 2021, item 11).

Act of 31 July 1981 on the remuneration of persons holding state managerial positions (Journal of Laws of 2020, item 1637).
Press Law Act of 26 January 1984 (consolidated text: Journal of Laws of 2018, item 1914).
Copyright and Related Rights Act of 4 February 1994 (consolidated text: Journal of Laws of 2020, item 288).
Constitution of the Republic of Poland of 2 April 1997 (Journal of Laws of 1997, No. 78, item 483).
Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, OJ L 167, 22.6.2001, p. 10.
Electronic Services Act of 18 July 2002 (consolidated text: Journal of Laws of 2020, item 344).
Treaty of Accession between the Kingdom of Belgium, the Kingdom of Denmark, the Federal Republic of Germany, the Hellenic Republic, the Kingdom of Spain, the French Republic, Ireland, the Italian Republic, the Grand Duchy of Luxembourg, the Kingdom of the Netherlands, the Republic of Austria, the Portuguese Republic, the Republic of Finland, the Kingdom of Sweden, the United Kingdom of Great Britain and Northern Ireland (Member States of the European Union) and the Czech Republic, the Republic of Estonia, the Republic of Cyprus, the Republic of Latvia, the Republic of Lithuania, the Republic of Hungary, the Republic of Malta, the Republic of Poland, the Republic of Slovenia, the Slovak Republic, concerning the accession of the Czech Republic, the Republic of Estonia, the Republic of Cyprus, the Republic of Latvia, the Republic of Lithuania, the Republic of Hungary, the Republic of Malta, the Republic of Poland, the Republic of Slovenia and the Slovak Republic to the European Union, signed in Athens on 16 April 2003, OJ L 236, 23.9.2003 (Journal of Laws of 2004, No. 90, item 864).
Act of 1 April 2004 amending the Copyright and Related Rights Act (Journal of Laws of 2004, No. 91, item 869).
Telecommunications Law Act of 16 July 2004 (consolidated text: Journal of Laws of 2021, item 576).
Act of 17 February 2005 on the computerization of the activities of entities performing public tasks (consolidated text: Journal of Laws of 2021, item 670).
Treaty on European Union and the Treaty on the Functioning of the European Union, OJ C 326, 26.10.2012, p. 1.
Personal Data Protection Act of 10 May 2018 (Journal of Laws of 2019, item 1781).

Case law

Judgement of the Court of Justice of the European Union of 8 September 2016 in case C-160/15 GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker (ECLI:EU:C:2016:644).
Judgement of the Court of Justice of the European Union of 26 April 2017 in case C-527/15 Stichting Brein v Jack Frederik Wullems, also trading under the name 'Filmspeler' (ECLI:EU:C:2017:300).
Judgement of the Court of Justice of the European Union of 14 June 2017 in case C-610/15 Stichting Brein v Ziggo BV and XS4ALL Internet BV (ECLI:EU:C: 2017:456).
Judgement of the Court of Justice of the European Union of 7 August 2018 in case C-161/17 Land Nordrhein-Westfalen v Dirk Renckhoff (ECLI:EU:C:2018:634).

Judgement of the Court of Justice of the European Union of 9 March 2021 in case C-392/19 VG Bild-Kunst v Stiftung Preußischer Kulturbesitz (ECLI:EU:C:2021:181).
Decision of the Court of Appeal in Warsaw of 21 June 2016, case ref. IA Cz 723/16.
Judgement of the District Court in Olsztyn of 6 June 2017, case ref. VII K 5/16.
Judgement of the Court of Appeal in Szczecin of 24 November 2016, case ref. I ACa 1159/15.
Judgement of the Supreme Court of 9 August 2019, case ref. II CSK 7/18.
Judgement of the Constitutional Court of 29 November 2017, case ref. P 9/15.
Judgement of the Constitutional Court of 6 December 2016, case ref. SK 7/15.
Judgement of the Constitutional Court of 10 February 2015, case ref. SK 50/13.
Judgement of the Constitutional Court of 12 December 2011, case ref. P 1/11.
Judgement of the Constitutional Court of 10 May 2005, case ref. SK 40/02.
Judgement of the Constitutional Court of 7 July 2003, case ref. SK 38/01.
Judgement of the Constitutional Court of 10 June 2003, case ref. SK 37/02.
Judgement of the Constitutional Court of 12 December 2011, case ref. P 1/11.
Judgement of the Court of Justice of the European Union of 17 October 2017 in case C-194/16 Bolagsupplysningen OÜ and Ingrid Ilsjan v Svensk Handel AB (ECLI:EU:C:2017:766).
Judgement of the Court of Justice of the European Union of 25 October 2011 in joined cases C-509/09 and C-161/10 eDate Advertising GmbH and others v X and Société MGN LIMITED (ECLI:EU:C:2011:685).
Judgement of the Constitutional Court of 12 May 2008, case ref. SK 43/05.
Judgement of the Constitutional Court of 14 December 2011, case ref. SK 42/09.
Judgement of the Constitutional Court of 20 February 2007, case ref. P 1/06.
Judgement of the Constitutional Court of 26 May 2008, case ref. SK 25/07.
Judgement of the Constitutional Court of 28 June 2000, case ref. K 34/99.
Judgement of the Constitutional Court of 29 September 2008, case ref. SK 52/05.
Judgement of the Constitutional Court of 30 October 2006, case ref. P 10/06.

Chapter III

The Regulation of Social Media Platforms in Hungary

András Koltay

1. Introduction

This overview will examine various issues related to the operation of social media platforms that have a significant impact on the public, along with their legal regulation and jurisprudence in the Hungarian legal system. The overview comprises two main parts. The first part will examine the general issues related to the definition of censorship and its application, as well as the various issues regarding the interpretation of censorship as it relates to social media and its various manifestations. The second part focuses on the legal means available to combat fake news and disinformation. The stakes are high for the public: According to the most recent data published by the Hungarian Advertising Association in 2020, the Internet had the largest share of the advertising market, with 41.45 per cent of the total advertising expenditure devoted to digital media, most of which was spent on various social media platforms.[1] According to research published by the Information Society Research Institute of the University of Public Service (Budapest), usage data show that Facebook—by far the biggest social media platform worldwide—also has an extremely strong market position in Hungary. While Facebook is used by 79 per cent of Internet users at least once a month (59 per cent use it daily and 16.5 per cent at least

1 *MRSZ Barométer–A válság hatásai a reklámiparban,* 2020.

András Koltay (2021) The Regulation of Social Media Platforms in Hungary. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 79–110. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_3

once a week), all the other social media platforms combined (Twitter, Instagram, LinkedIn, etc.) are used only by 27 per cent of users at least once a month (12 per cent daily, 11 per cent at least once a week).[2] Legal regulation is key to guaranteeing the proper functioning of the public sphere while enhancing democracy. However, this overview is intended only to present existing legal approaches in regulation and practice and not to propose future regulatory directions.

2. Content regulation on social media platforms

Social media platforms are subject, like other media, to general legislation that restricts freedom of speech and the public expression of opinions. In addition to these laws, other special rules affecting the content of opinions may also apply, which are adjusted to the unique characteristics of the medium concerned. Legal regimes are forced to apply unique solutions to the problems raised by social media platforms that are not applied in the regulation of legacy media.

2.1. Prohibition of censorship

The principle that censorship of specific content is not permissible constitutes a cornerstone of freedom of the press. Obviously, it is not entirely clear what exactly is meant by censorship. A narrow interpretation of the term 'censorship' might be construed to mean restrictions that are arbitrary without legal or judicial safeguards (which can be anticipated), hence making publication impossible. In addition to this, in public discourse, the term 'censorship' is also used to refer to post-publication restrictions applied after disclosure or publication. Originally the word 'censorship' referred to state interference with the content published by the media. As the notion of freedom of the press has developed over time, it has become a generally accepted view that 'censorship,' as arbitrary intervention in content, is impermissible, whereas *a posteriori* accountability or prosecution for the publication of unlawful content may be acceptable.

According to Frederick Schauer, the meaning of the term 'censorship' has become somewhat hazy.[3] On the one hand, censorship can be carried out not only by the state, but also as a result of various processes in society. While censorship can obviously still come from the state, it can also come from private companies. Censorship can be either direct or indirect in nature. Another type of censorship occurs when one person or more chooses to stay out of the public eye in response to an opinion against them published by others, for example, as a result of hate speech; this is called the silencing effect.

2 *Bizalom, tudatosság, veszélyérzet az interneten,* 2020.
3 Schauer, 1998.

Following Schauer, we can conclude that an overly broad application of the concept of censorship can render its scope vague, thereby devaluing it; in this way, censorship will no longer necessarily be as serious a threat as it was for the Hungarian revolutionaries of the mid-nineteenth century (contesting the Habsburg reign), for example. As a result, certain content subject to censorship may be left without proper protection, even when the highest level of protection would be justified. Similarly, the term must be used carefully when discussing censorship by or with the participation of social media platforms.

2.1.1. Censorship in Hungarian history

In Hungary, the abolition of censorship was among the first of the twelve demands of the youth of Pest at the beginning of the revolution against Austrian imperial power in March 1848. More than half a century prior to the revolution, at the end of the eighteenth century, after Francis I, Emperor of Austria and King of Hungary ascended to the throne, the rules of press censorship had become significantly harsher. The development of newspapers and periodicals as a result of the Hungarian Enlightenment abruptly slowed down with the gradually increasing rigor of the censors from 1793 onward and was later paralyzed by the suppression of the Jacobin movement. The emperor's decree of 25 June 1793 made inspecting books and licensing the operation of printing presses a royal right, and by 1795, all periodicals had closed, the number of newspapers had fallen sharply, and their content had become anodyne. A royal charter of 1806 further required a royal patent to open a bookshop. The turnover of booksellers and lending libraries was also monitored separately on the basis of a chancellery order issued on 5 June 1818. Two types of censorship were developed: revision and actual censorship.[4] Revision—regulated by a court decree of 18 April 1793—entailed the oversight of books and press products imported from abroad. Only books, periodicals, and newspapers approved by the central book inspection agency were allowed to be imported across the border. Censorship—as regulated by the decree of 25 February 1795—involved prior licensing by an appointed official censor, whereas its counterpart was the post-publication inspection of the publications submitted as legal deposit copies. Pursuant to the decree of 18 April 1793, each and every printing press was obliged to hand over three copies of each of their publications to the revisor. The publications were read through and, if any objections were raised, the censor in charge was held accountable, and this made them exercise the utmost caution in their work.

A provision of the Press Act (Act II of 1986)—repealed in 2011—allowing prior restraint was effective until 1997 in Hungary. Pursuant to Article 15(3) of the Act, exercising freedom of the press would have constituted a criminal offense or an incitement to commit such an offense, in the event it would have caused a breach of public morality or of someone's moral rights. Moreover, if the newspaper had been

4 Bényei, 1994, pp. 15–17.

distributed before registration under the requirement of notification, the court could prohibit the 'public communication' of the press product concerned on the motion of the prosecutor. The prosecutor had the right to suspend publication temporarily until the court reached a decision.

According to a motion submitted to the Hungarian Constitutional Court (CC) in 1997, this rule was unconstitutional in restricting the freedom of the press. The CC found the provision partly unconstitutional. However, since the constitutional and unconstitutional content were included in the very same sentence—and the CC 'has no right to rewrite the law'—the CC decided to annul the entire contested provision (20/1997. (III. 19.) AB). The decision did not examine the provision primarily from the point of view of the freedom of the press, thereby implicitly acknowledging that prior restraint is not incompatible with the fundamental right. Most members of the CC were of the opinion that the provision of the law, according to which the prosecutor had the right to request the prohibition of publication in the case of a violation of the moral rights of others or in private prosecutions of crimes—regardless of the will of the victims—violated the right to self-determination. By contrast, the prior restraint on the grounds of public morals was not found to be unconstitutional with regard to criminal offenses, subject to public prosecution or in the event of failure to meet the requirement of notification.

Clearly, the concept of censorship is historically linked to the state, a potentially oppressive mechanism capable of acting against the freedom of speech. However, in the modern media world, since the second half of the twentieth century, the scope of the concept has grown considerably, and censorship as a legal concept is used much more widely than before. On the one hand, censorship is no longer used only in relation to state restrictions, as various private interests (such as advertisers) are also able to restrict media content; on the other hand, censorship is not necessarily applied as a result of external pressure, which recognizes the possibility of internal so-called self-censorship. It is conceivable that the publication of certain specific content may also be required by law, such as communications published as a result of the exercise of the right of reply or public service media providers' obligation to publish political advertisements during election periods.

2.1.2. Censorship and social media

With the rise of social media, platform operators have emerged as players capable of imposing restrictions on the content made available to the general public. These service providers have a number of means of restricting freedom of speech, either through service settings (algorithms or moderators) and instructions or through case-by-case decisions about specific content. Service providers may interfere with others' rights to free speech to further their own business or political interests, or in cooperation with some oppressive state regimes.

However, it is important to underline that, in a strict legal sense, intervention by social networking website operators in the communication process (compiling a search engine results list or the social media platform feed, which will necessarily

suppress some opinions by deleting some links or content from the service or moderating comments) cannot be considered censorship, even within the broader interpretation of the term referred to above. Instead, it can be seen as the exercise of rights derived from private property and other subjective rights, which—in the absence of statutory requirements—are free and not prohibited if they use legal means, even if they may be morally objectionable, for instance.

Censorship is traditionally understood as public authorities' arbitrary interference with the exercise of freedom of the press. In such cases, platforms can become the facilitators of public authorities. It follows from the European Union (EU) Regulations, among others, that platforms are required to decide whether specific user content is infringing or not. In certain situations they are obliged to do so by law.[5] In democratic states, and hence in the EU member states, this state interference cannot be regarded as censorship, but it is clear that platforms' content decisions are completely lacking in safeguards for the protection of the fundamental right.

2.2. Regulation of social media

Social media platforms in Europe are primarily regulated by common EU rules. Member states have room to maneuver on regulation, but it is narrower. When reviewing regulatory issues regarding platforms, one should not forget the regulation created and overseen by the platforms themselves, which is known as private regulation, as distinguished from legal regulation.

2.2.1. Regulation imposed by the European Union

If the gatekeepers merely provide a technical service by making available, storing, or transmitting others' content (similarly to printing houses or news agents), then as long as they are not aware of any infringements, they should not be held liable for infringements committed by others. According to the European approach, after they have become aware of an infringement, they can be held liable for their own failure (that is, for the failure to take down the infringing content). The EU Directive on Electronic Commerce, which aims to regulate this issue, imposes the obligation of takedown after becoming aware of an infringement on certain intermediary service providers.[6]

The activities of gatekeepers covered by the Directive—simple transmission, caching, and hosting—play an important role in online communications, but since the legislation was adopted in 2000, legal liability has arisen for a number of gatekeepers that either did not exist at the time or which were not covered for other reasons, such as search engines or social media. In the absence of a better solution, the courts apply various analogies to these gatekeepers, such as by classifying some of them as hosting providers.

5 Directive 2000/31/EC, art. 14.
6 Directive 2000/31/EC, arts. 12–14.

The material scope of the legislation is important: the EU Directive grants gatekeepers an exemption from liability, even if they transmit infringing content, provided certain conditions are met. This system of exemptions itself has not necessarily become obsolete, but one thing has certainly changed since 2000: Today's gatekeepers are increasingly less likely to be considered actors that merely store or transmit data and are passive with regard to its content; although the content is still produced by their users or by other actors independent of them, the services themselves select, arrange, prioritize, de-prioritize, delete, or make inaccessible the content in their systems. An equitable rule in the Directive exempts the passive actor until it becomes involved (i.e., until it becomes aware of the infringement), but it seems that this is not the only conceivable approach in this respect for the new types of gatekeepers. Although it is still true that the volume of content these new gatekeepers manage makes the pre-publication monitoring obligation both impossible and unreasonable, the same is true of comprehensive post-publication checking without an external call for attention.

Articles 12 to 14 of the Directive grant broad exemption to intermediary service providers. For hosting service providers, this means that if the content they transmitted or stored was not their own, and if they were not aware of the infringing nature of such content, they are not held liable, provided that they take immediate action to take down the content or terminate access to it (Article 14). However, failure to do this could result in the service provider being held liable for this as if it were its own infringement. In addition, the Directive also stipulates that no general infringement monitoring obligation can be imposed on intermediaries (Article 15). This general prohibition appears to have been undermined by the judgment of the Court of Justice of the European Union (CEJA) in *Glawischnig-Piesczek v. Facebook*,[7] in which the CJEU ruled that it was not contrary to EU law to oblige a platform provider to delete posts with similar or the same content as a defamatory post that has previously been declared unlawful.

The Commission's subsequent Recommendation reinforces this approach, including the exceptions to liability (notice-and-takedown system) set out in the E-Commerce Directive as a solid basis for dealing with illegal online content.[8] Although the document applies to states, it aims to broaden gatekeeper obligations and responsibilities through state legislation relating to notification and proper processing of user requests, the possibility of counter-notification by hosting providers, transparency (which seems to be the magic word in these disputes), and procedural safeguards. More importantly, however, "hosting service providers should be encouraged to take, where appropriate, proportionate and specific proactive measures in respect of illegal content,"[9] but "there should be effective and appropriate safeguards to ensure that hosting service providers act in a diligent and proportionate manner in

7 *Eva Glawischnig-Piesczek v. Facebook Ireland Ltd.*
8 Commission Recommendation (EU) 2018/334.
9 Commission Recommendation (EU) 2018/334, s. 18.

respect of content that they store."[10] The Recommendation clearly demonstrates the Commission's approach of strengthening the regulatory mechanisms already in place by formalizing gatekeepers' (hosting service providers') existing non-legal procedures and policies.

In addition to the E-Commerce Directive, several more general pieces of legislation also apply to communications via social media platforms, including laws on data protection, copyright, protection of personality rights, public order, and criminal law. Such legal provisions may also introduce special obligations for hosting service providers in the context of taking down violating content.

Offline restrictions on speech are also applicable to communications through social media platforms.[11] Common violating behaviors on social media can be fitted into a more traditional criminal category (that is, one that was adopted in the context of the offline world) almost without exception, making the introduction of new prohibitions unnecessary.[12] However, this duality gives rise to numerous difficulties, as, on the one hand, such limitations are defined as part of the national legislation of each and every country (and the law of free speech is also far from being fully harmonized among EU member states), and, on the other hand, social media constitute a global phenomenon by nature, meaning that it transcends national borders. For instance, an opinion that is protected by the freedom of speech in Europe might constitute punishable blasphemy in an Islamic country. Since harmful content can be made available worldwide and shared on a social media platform quickly, the absence of a uniform standard can lead to tensions and even violence.[13]

On-demand media services that can also be accessed through the Internet have been subject to the scope of the Audiovisual Media Services (AVMS) Directive since 2007,[14] but social media are not counted among such services. The main reason for this is that providers of on-demand media services bear editorial responsibility for the content they publish; they order and purchase such content, and they have a final say in publishing a piece of content.[15] In contrast, social media operators only provide a communication platform, but may not make any decision regarding a piece of content before it is published (the situation is different if some kind of preliminary filtering is used, but such filtering only relates to specific categories of content). As social media platforms spread, it became clear, about a decade after the previous amendment of the Directive, that media regulation cannot be interpreted in such a restrictive manner any longer. To address this, the AVMS Directive introduced the terms 'video-sharing platform service' and 'video-sharing platform provider.'[16] Even though the original proposal would not have extended the Directive's scope to social

10 Commission Recommendation (EU) 2018/334, s. 19.
11 Rowbottom, 2012, pp. 357–366.
12 House of Lords, 2014.
13 Kohl, 2018.
14 Directive 2010/13/EU.
15 Directive 2010/13/EU, art. 1.
16 Directive 2010/13/EU, art. 1(1)aa.

media platforms in general (as it applied to the audiovisual content uploaded to such sites), it became clear during the legislative process that they could not be exempted from the Directive by focusing on portals used primarily and actually to share videos (such as YouTube).[17] This means that despite its somewhat misleading name, a video-sharing platform also includes social media where audiovisual content is published. An important aspect of the newly defined term is that service providers do not bear any editorial responsibility for such content; while service providers do sort, display, label, and organize such content as part of their activities, they do not become media service providers.

Article 28b of the amended Directive provides that Articles 12 to 15 of the E-Commerce Directive (in particular the provisions on hosting service providers and the prohibition of introducing a general monitoring obligation) remain applicable. In addition to this, member states must ensure that video-sharing platform providers operating within their respective jurisdiction take appropriate measures to protect children, combat hate speech and content in support of terrorism, and comply with the rules relating to commercial communications.[18]

2.2.2. Provisions of the Hungarian legal system

At the outset, it is worth noting that in the Hungarian legal system, while the activities of social media platforms are regulated by law, judicial practice is very fragmented, which may be due to the difficulties of enforcement against the largest service providers. The Hungarian regulation on electronic commerce services[19] was developed by implementing the E-Commerce Directive in 2001. Based on the 'country of origin' principle, the scope of the E-Commerce Act covers information society-related services provided to and from Hungary, the providers of such services, and the recipients of such services.

The Hungarian legislation—in line with the E-Commerce Directive—lays down, as a general rule, the liability of intermediary service providers for the information they make available to the public, while also specifying the cases in which their exemption from this liability is guaranteed. It should be highlighted that this regulation covers liability under civil law, criminal law, and public administration law alike (and also the possibility of exemption). Assuming that intermediary service providers' activities only include information storage, transmission, and making it available, they hence cannot be obliged to monitor information or to identify circumstances that indicate unlawful activities. As a consequence, the liability for the information produced and published by a content provider on the Internet is direct,

17 Robinson, 2017.
18 Directive 2010/13/EU, art. 28b(3).
19 Act CVIII of 2001(hereinafter 'E-Commerce Act').

while the liability of the intermediary service provider, which is only a passive actor in the content production process, is limited.[20]

The thinking behind the provisions of the E-Commerce Act governing the scope of exemptions from liability for intermediary service providers regarding the information they store, transmit, or make available is based on the fact that, on the one hand, service providers are not liable if their activity is purely technical, automatic, and passive, involving the transmission of information to the public (or to the recipient), while on the other hand, once they become aware of the unlawful nature of the content, they must take immediate action to take it down (Articles 8 to 11).

The purpose of the notice-and-takedown procedure under the E-Commerce Act is to offer the affected parties an alternative to lengthy and cumbersome court proceedings to establish the infringement and remedy the infringing situation by giving right-holders the possibility to restrict access to the infringing information and remove the infringing content. The legislation also regulates in detail the process and conditions of the notice-and-takedown procedures as they relate to copyright infringements and the takedown of content infringing the personality rights of minors. It should be noted that this form of procedure merely prevents, but does not exclude, the possibility of bringing a claim before the courts, and that the relevant rules only apply in the relationship between the service provider and the injured party, not in any court proceedings (Article 13).

In Hungary, the implementation of the provisions of the AVMS Directive on video-sharing platforms was achieved through the amendment of the E-Commerce Act—in the course of which the legislator mainly adopted the Directive's rules, but also imposed additional obligations on video-sharing platforms through certain detailed rules. The amendment obliges video-sharing platform providers to take appropriate measures and apply suitable technical solutions if the content of their service endangers the physical, mental, spiritual, or moral development of minors; offends human dignity or constitutes hate speech, a criminal offense, or an incitement to commit such an offense; or violates the rules on commercial communications (Article 15/D(1)). Similarly to the AVMS Directive, the Hungarian legislation does not specify such appropriate measures, but lets the service providers determine these for themselves. In order to protect minors, the law provides for the use of age verification and parental control systems (Articles 15/F-H).

In addition to this, with regard to commercial communications, the E-Commerce Act stipulates the application of the rules specified in the media regulations—Article 20(1)-(7) of the Act on the Freedom of the Press and the Fundamental Rules of Media Content (Press Freedom Act)[21] and Article 24 of the Act on Media Services and Mass Media (Media Act)[22] which means that video-sharing platform providers in Hungary must comply with the same obligations as media service providers with regard to the

20 See: court decision BDT2008.1777.

21 Act CIV of 2010 (hereinafter, 'Press Freedom Act').

22 Act CLXXXV of 2010 (hereinafter, 'Media Act').

organized commercial communications that they distribute or sell. The Media Act and the Press Freedom Act—in addition to the definition of video-sharing platform services and the clarification of the registration rules—do not contain any material requirements for social media platforms.

2.2.3. Private regulation of social media

'Private regulation' refers to a system in which the platforms themselves create rules and oversee them in a process that they also create themselves. These rules do not, of course, oblige the platform itself, but in the first place, its users, although platforms may also be obliged to work within the system (though it cannot be legally enforced against them). Private regulation is thus the additional extralegal regulation of user behavior, which may overlap with codified legal regulation, but which is not a necessary feature of it. Platforms may enforce private regulation on their users through their contract with them, so these rules have legal binding force between the parties. Furthermore, because it primarily concerns content that may be published and shared by users, it directly affects freedom of expression. The Oversight Board[23] established by Facebook can also be considered private regulation, even if the social media platform tries to suggest in its communication that this Board is independent from it. The Board's rules of operation are established by Facebook; its members are invited by Facebook and its competence extends exclusively to the Facebook platform. The establishment of the Oversight Board is another step toward the construction of a 'pseudo' legal system that develops in parallel with the state legal system.

Platforms have the right to create these rules, which stems from their right to property and the right to freedom of enterprise. There are relatively few restrictions on such private regulation, although platforms are required to comply with restrictions on freedom of speech (for instance, with regard to the advertisements they may accept) and with the requirement of equal treatment of their users. As such, in addition to protecting their rights, private regulation may impose other restrictions on the opinions published on the platform. Jack Balkin calls this phenomenon 'private governance,'[24] while others prefer to use the less euphemistic term 'private censorship.'[25] As Balkin warns, it seems unreasonable to attempt to discuss compliance with government regulations separately from private regulation, considering that the threat of government regulation incentivizes platform providers to introduce private regulations because the providers have an interest in avoiding any troublesome government interference.[26]

23 See the website of the body: https://oversightboard.com.
24 Balkin, 2018, pp. 1179 and 1182.
25 Heins, 2014.
26 Balkin, 2018, p. 1193.

The removal of content that is undesirable for the platform concerned is not the only means of implementing private regulation. A far more significant means is the editing and sorting of content presented to individual users, as well as the promotion and suppression of certain pieces of content, the impact of which is not limited to individual pieces of content but to the entire flow of content on the platform. This does not constitute 'regulation' because it does not require a normative decision on the 'suitability' of the content (examined in the light of the private regulation code), but it fundamentally affects the chances of a piece of content reaching the public, and so it may be regarded as a kind of editing that has more impact on the fate of each piece of content than private regulation.

The enforcement of the freedom of speech on a social media platform is much more dependent on the rules applied and implemented by the various platforms than on the government (legal) regulations relating to freedom of speech as a fundamental right. The standards, policies, and service terms and conditions social media platforms apply result in decisions made in bulk, and they cannot be matched by any lengthy legal proceedings that might be launched in individual cases. In addition to platform ownership, a contract by and between the platform and each of its users serves as the legal basis for the platform's capacity to interfere with its users' freedom of speech. The platform determines the provisions of that contract. Users are not in a position to request the amendment of the contract, while it may be amended by the platform unilaterally at any time. It is also important to note that the same contract is concluded with each and every user. Even though the contract and the interference permitted by it affect the exercise of a constitutional right, and countless debates, conversations, and exchanges of information are taking place on the platform at any given time in connection with public affairs, no interference by the platform can be considered state action, and the platform itself is not considered a public forum. An action taken by a platform, even if it restricts its users' opinions, cannot be attributed to the government, and as such it is not subject to any constitutional safeguard relating to the freedom of speech.[27]

When there is a conflict of interest or a dispute between the platform provider and the user affecting the exercise of freedom of expression we must hence, in a somewhat sobering way, seek a solution in contract law instead of in constitutional doctrines.[28] When a user decides to subscribe to a platform and accepts that platform's terms and conditions by a simple mouse click, they become subject to 'private regulation,' including all content-related provisions, and the safeguards of free speech are no longer applicable to the user's relationship with the platform.[29] It should not be a surprise that the contracts all the major platforms use are carefully considered and precisely drafted documents (or that they knowingly use vague language in order to extend the platform's discretionary powers).

27 Fradette, 2013–2014, pp. 953–957.
28 Fradette, 2013–2014, p. 971.
29 Fradette, 2013–2014, p. 977.

However, consumer protection does not seem to provide much opportunity to protect users' freedom of speech when a rule set by a platform and its application seem reasonable and justifiable, and not arbitrary. Indeed, they typically are; even though they might be questionable, this is not evidence of any violation of the consumers' rights in and of itself. It also seems difficult to object to the application of such policies on a legal basis, considering that a platform is free to determine its own policies and instruct its moderators without being required to respect the constitutional safeguards and legal limitations of freedom of speech. A user's only option is to show that the platform removed a piece of content it was not authorized to remove,[30] something that seems well-nigh impossible to demonstrate due to the widely defined limitations of content and the platform's broad discretionary powers. A user may also try to make use of the existing anti-discrimination rules if their right to equal treatment has been violated, but proving a breach in such a situation (showing that a piece of content was removed which was not removed when published by another user) seems rather difficult, and the enormous volume of content coupled with the absence of a monitoring obligation on the platform's side (which the platform may invoke as a defense) also considerably limit the user's chances.

2.3. Censorship and social media

Social media platforms take part in the supervision of the content published by users through the enforcement of national (or EU) legislation on the one hand and through the means of private regulation created by them on the other hand. An important difference between the two is that while in the first case, the platforms are legally obliged to take down certain content, in the second case, they decide to do so on the basis of a voluntary initiative.

2.3.1. Possibilities for taking down content based on law

While participation in co-regulatory schemes is generally voluntary for service providers, on a single important issue, the law still subjects them to specific co-regulation at the European level. This implementation of co-regulation obliges platforms to participate in monitoring the legality of user behavior. The regulation is binding on the platform, but it aims to take action against infringements committed by users. The platform's liability is not for publishing infringing content, but for failing to take action against it.

Article 14 of the E-Commerce Directive—which is followed by the national legislation—provides for a broad exemption for platforms. Hence, if they did not make their own content available and were not originally aware of the infringing nature of that content, they will not be held liable, as long as upon becoming aware of its infringing nature, they take action to take down or terminate access to the infringing

30 Fradette, 2013–2014, p. 957.

content without delay. In the event of failure to do so, however, they may be held liable for their own omission. In this way, codified legal regulation forces the platforms into a decision-making role with regard to user content, expecting them to make a decision on the illegality of the content, conditional upon them becoming aware of it. The consequence of this procedure may be the takedown (removal) of the content.

The assessment of the 'infringing' nature of content raises a very important issue. The takedown obligation is independent of any judicial or other official procedure to establish the infringement, and the hosting provider must act before such a decision is made, if any legal proceedings are instituted at all. It is therefore up to it to decide on the infringement itself, and this decision will be free from the guarantees of the rule of law (while it may also affect the freedom of expression) and will encourage the obligee to decide against the preservation of the content in case of any possible concerns in order to save itself. This co-regulation, enforced by legal regulation, may be seen as a specific form in which the enforcement of codified legal norms (restrictions on freedom of expression) is monitored by a private party (the platform), while at the same time enforcing the sanction (deletion of content).

In the Hungarian case law on social networking websites, users typically do not attempt to enforce the notice-and-takedown procedural obligations of the E-Commerce Act—introduced in the Hungarian legal environment as a result of the E-Commerce Directive—before the Hungarian courts; users of social networking websites instead typically try to settle disputes between themselves through traditional judicial channels.

2.3.2. The application of private regulation and the restriction of debates on public affairs

Another way of taking down user content is through the enforcement of platforms' private regulation. Unfortunately, content moderation that transgresses the legal boundaries of freedom of speech is a very common and highly criticized means employed by platforms.

Several Hungarian cases have arisen related to the private regulation of social media platforms. A video posted by Minister János Lázár on Facebook in March 2018 during the campaign period for the Hungarian Parliamentary elections made headlines. The footage shot on the streets of Vienna showed the Austrian capital as a dirty, unsafe, and less livable city with the implication that this was due to the large number of immigrants. The platform took down the video on the grounds of violation of the rules of the community standards applicable for the prohibition of hate speech. Following a complaint submitted by Lázár, Facebook ultimately made the video accessible again on the politician's social media page. The official reasoning behind the decision was that the topic the minister addressed (that is, immigration) had significant news value and public relevance; this circumstance constitutes an

exception for breaches of the rules on hate speech and justifies a reversal of Facebook's previous decision on the breach of community standards.[31]

Even more serious restrictions were imposed on the president of the Mi Hazánk Mozgalom Party, László Toroczkai, and then on the party itself. First, during the final phase of its campaign for the 2019 European Parliament elections, the party leader's profile, with more than 200,000 followers, disappeared without any prior warning or notice.[32] The president of the party, on the grounds of what he perceived as politically motivated censorship, eventually filed a personality right lawsuit against the platform, claiming damages for the alleged harm he suffered.[33] A year and a half later, in October 2020, Facebook deleted a page advertising an event the party organized to commemorate the 1956 Revolution, and a few days later, Mi Hazánk's official social media page, also citing a violation of the community standards,[34] but the party did not receive a more detailed justification from the platform for the decision.[35]

A case that occurred in February 2020, when Google deleted news portal Pesti Srácok's YouTube channel, which had been operating for five years, along with all its content and without any prior warning, is also an example of censorship by social media platforms. The case arose from a video the news portal uploaded to expose an alleged pedophile offense, in which elements of another video—not considered by the journalists to be of concern—were edited in as illustrations, thereby violating the community rule prohibiting the depiction of the sexual abuse of children. In this regard, it should be underlined that, on the one hand, the scenes in the video that could be considered potentially questionable were obscured, and on the other hand, the news portal claimed that the video was not made public; it was only saved as a draft on the editorial interface—despite this, the channel was deleted.[36] According to the operator's response, the channel was deleted due to a breach of the rules prohibiting the depiction of the sexual abuse of children.[37] As a consequence of the case, two other channels also registered by the same editorial staff suffered the same fate in the days following the incident, but the reasoning was not even attached to the decisions in these cases.[38]

In addition to removing individual user content or even deleting entire user accounts, a common solution social networking site operators use is the so-called shadow ban—a way of restricting access to individual content and user profiles and reducing their visibility in a less obvious and noticeable way, without the person concerned or their account's friends and followers being aware of it. Platforms rarely mention the use of this tool, yet user experience and reports indicate that the reach of some users' posts to the general public has been significantly reduced. Hungary's

31 Gilbert, 2018.
32 *Toroczkai törlése után megregulázná,* 2019.
33 *Beperli a Facebookot a Mi Hazánk elnöke,* 2020.
34 *Törölte a Mi Hazánk oldalát,* 2020.
35 Pálfy, 2020.
36 *Ez már hajtóvadászat,* 2020.
37 *Gyerekpornós részlet miatt törölték,* 2020b.
38 *Ez már hajtóvadászat,* 2020.

Minister of Justice Judit Varga suspected such interference with her own account in January 2021.[39] According to Facebook, this type of restriction applies to content that does not have to be removed based on community standards, but which is still considered problematic.[40]

2.4. Legal disputes between users

Disputes between users, with charges pressed based on claims of defamation, libel, or slander, must be judged according to the provisions of the Hungarian Civil Code or Criminal Code in the same way as if the alleged infringement had been committed in the legacy media.

When civil proceedings are brought to remove infringing content, the court may order the publisher of the infringing content to cease the act, prohibit the publisher from engaging in the activity, or order the publisher to terminate the infringing situation as a sanction for the infringement of the personality right. However, these measures are not imposed on the hosting service provider to take down the content, but instead oblige the person who has actually committed the infringement to perform a certain act. A good example of this is decision no. BDT2019.3987 of the Szeged Regional Court of Appeal, issued regarding a user's blatantly obscene comment about the mayor of a municipality. The plaintiff brought a defamation action directly against the defendant, as a result of which the court, in addition to finding a violation of the right to honor, prohibited the defendant from using any further defamatory language against the plaintiff.

A substantial body of court jurisprudence has grown up in Hungary in relation to comments posted on online content. The main issue in the debate is whether the rules of the E-Commerce Act or those of the Civil Code apply to the liability of platforms that provide the opportunity to comment; in other words, whether, upon learning of an infringing comment (infringing personality rights), prompt removal may lead to an exemption (pursuant to the E-Commerce Act) or the involvement in the publication of the infringing content would automatically lead to the establishment of strict liability (in accordance with the Civil Code). The Pécs Court of Appeal stated in its decision BDT2013.2904 that:

> The operator (intermediary service provider) of the website accessible via the Internet is subject to civil law liability for the content of comments that violate reputation, uploaded by others to the website, if it does not initiate the removal (takedown) without delay upon gaining knowledge of it.

39 Varga, 2021.
40 *Remove, Reduce, Inform,* 2019.

However, the Hungarian Curia and the CC considered (based on the Civil Code's approach) the provision of an option to comment to be a unique type of dissemination (it is not the platform itself that publishes the infringing post, but it becomes involved in the process as a result of the publication, and therefore courts consider this to be 'dissemination,' which triggers the same civil law liability for defamatory statements as publication itself).

According to the decision of the Constitutional Court of Hungary 19/2014. (V. 30.) AB, a website operator may be held liable for published comments with infringing content even if it does not moderate the remarks because the rights of the person affected by the infringing communication could not be protected if the operator of the Internet site did not have liability for the infringing comments, since liability for the infringing communication is based solely on the fact of infringement.[41] The single decision related to online comments which the CC has taken so far has been subjected to considerable criticism.[42] In that case, the fact of infringement was not disputed between the parties, nor was it disputed by the website operator (therefore, the CC did not conduct a substantive investigation in this respect), but instead, relying on the E-Commerce Act, it merely objected to the fact that it would be subject to any liability for 'alien' content, regardless of whether the entry had been taken down (removed) without delay upon notification (Paragraph [56]). The CC found that the website operator's liability is based solely on the fact of the infringement; other circumstances cannot be decisive. In the event of an infringement, no distinction can be made between different websites and the operator's legal liability (Paragraph [64]).

The statement of reasons for the decision is contradictory in many respects and unfortunately fails to provide a general approach applicable to the constitutional characteristics of the online public sphere, which functions according to a different logic from the legacy media and which is increasingly important for the discussion of public affairs.[43] Nevertheless, it should be noted that as far as liability for infringing comments is concerned (which was the most crucial issue to be settled by the decision), the approach applied by the CC is not without precedent: It was in harmony with the Hungarian jurisprudence prevailing at that point in time. Furthermore, the European Court of Human Rights (ECtHR) did not subsequently find such application of the general civil law liability rules on the grounds of violation of the personality rights to be incompatible with Article 10 of the European Convention on Human Rights (ECHR) either. This was the way in which the ECtHR had reached decisions prior to the CC decision in *Delfi v. Estonia* and also in another subsequent case that was filed with the ECtHR as a follow-up case based on the case underlying the CC decision.

41 Koltay, 2015.

42 See, for instance: Grad-Gyenge, 2015; Klein, 2016.

43 For a more detailed and informal explanation of the statement of reasons, see the text written by the judge-rapporteur in the case, Bragyova, 2016.

In 2016, the Fourth Section of the ECtHR awarded in favor of the applicants in *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt. v. Hungary*.[44] The Strasbourg Court found the decision of the Hungarian courts to be in breach of Article 10 of the ECHR, mainly due to the less offensive nature of the comments compared to the comments in *Delfi*.[45] The *Magyar Tartalomszolgáltatók Egyesülete* judgment ruled that the subject of the original article (to which comments were posted) concerned a public matter,[46] and this is of material importance. Increased protection of freedom of expression is essential for the discussion of matters of public interest.[47] The ECtHR considered not only the articles concerned but also the comments made on them to be defensible due to their participation in a debate on public affairs, thus according them outstanding importance.[48] In *Delfi* and *Magyar Tartalomszolgáltatók Egyesülete*, the target of the comments was a legal entity, the operation and criticism of which clearly qualified as a public matter. In any case, the protection of a legal entity's rights should be assessed in a different way to that of natural persons; the 'moral' rights of legal entities deserve less strict protection[49]—and the ECtHR considers it necessary to make this distinction.[50] The ECtHR also considered the consequences for content providers to be serious, even without an obligation to pay compensation. Although the Hungarian courts, in addition to establishing the infringement, also ordered payment of the not excessively high court fees, according to the ECtHR, the wider consequences of the decision must also be taken into account, for example, the incentive for them to close down their interfaces to commenting in the future and to refrain from providing their readers with this option.[51] The reason for Hungary's censure was not the application of civil liability *per se,* but a different assessment of the content of the comments, which—according to the ECtHR—could not have been considered to be infringing. The approach taken by the CC on the main issue was therefore compatible with Strasbourg's practice relating to Article 10 of the ECHR.[52]

The case law is also developing on non-anonymous comments published on Facebook. In these cases, although the identity of the commenter is known, the plaintiffs sued the users who allowed the comment (under whose post the offensive comments were published). The Curia applied its practice on anonymous commenting in decision BH2016.330: "The holder of a Facebook profile is liable for the communications of unlawful content appearing on his profile page." In contrast, in decision

44 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt. v. Hungary*.
45 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt. v. Hungary*, paras. 76 and 91.
46 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt. v. Hungary*, para. 72.
47 Barendt, 2005, pp. 155–162.
48 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt. v. Hungary*, para. 72.
49 See also: *Uj v. Hungary*, para. 22.
50 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt v. Hungary*, para. 65.
51 *Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt v. Hungary*, para. 86.
52 With regards to the assessment of the decision of the *Magyar Tartalomszolgáltatók Egyesülete*, see: Sepsi, 2015; Szigeti, Simon, 2016.

BDT2017.3675, the Pécs Court of Appeal applied the requirements of the notice-and-takedown procedure, but apparently by applying the Civil Code:

> The content provider fulfils the requirement of conduct that can be expected under the circumstances where, in the event of a complaint against a comment posted to a content published on its Internet portal (Facebook site), it removes that comment without delay.

The decision does not refer to the E-Commerce Act, while the Civil Code does not acknowledge the notice-and-takedown procedure, i.e., gaining knowledge of the infringing nature of a piece of content as a condition for establishing liability or takedown as an option leading to release from liability.

3. The possible regulation of fake news

Ever since the 2016 American presidential election campaign, the term 'fake news' has become a popular way to describe a form of propaganda that deliberately spreads false information as widely as possible, mainly through online platforms and social media. Probably the most widely known example of fake news is a scandal commonly called 'Pizzagate,' referring to rumors that Democratic candidate Hillary Clinton was operating a child trafficking network in a Washington pizzeria. The owner and staff of the restaurant were sent death threats, and one person even entered the restaurant and fired a gun to rescue the children.[53] It was also revealed that a group known as the 'Macedonian teenagers' made their living in Veles, a poor Macedonian town, by producing and spreading fake news aimed at the American public.[54]

3.1. The concept and problems of fake news

Lies in the media are, of course, by no means a recent phenomenon.[55] Various ancient, mediaeval, and modern examples are known portraying the deliberate dissemination of false information,[56] and this is one reason why only fragmentary knowledge can be obtained about human history. However, in the age of the Internet and social media in particular, the volume of this information and the speed at which it spreads constitute a new development that is changing the quality of the public sphere. For the time being, there appear to be no effective legal means of combatting

53 Robb, 2017.
54 Subramanian, 2017.
55 Bernal, 2018, pp. 230–234.
56 Darnton, 2017.

the spreading of fake news via online platforms. On these platforms, the competition is for the audience's attention, for the seconds a user spends on a piece of content,[57] so it is necessary to produce as exciting, interesting, and viral a piece of content as possible, even if it means lying.

The communication that takes places on these platforms has a fundamental impact on political culture and democratic processes in general, an impact which is negative in many respects, not only because of the ease with which lies can be spread, but also because it has led to the increasing superficiality of public debates. For economic reasons, platform providers are interested in creating public spaces that are as intensive—and thereby as superficial—as possible.[58] However, spreading lies *en masse* does not serve the interests of a democratic public sphere. It is also questionable whether this disregard for the truth is compatible with the traditional philosophical underpinnings of free speech. If one regards freedom of speech as a tool of community-based democratic decision making, then the deliberate publication of lies does not serve that purpose. At the same time, technological developments have spawned a new generation of lies, such as deepfake videos, where the face of a person in a real recording is replaced by someone else's face, thereby creating the false impression that the latter was saying or doing something, although they actually did not. In this way, any words can be put into the mouths of public figures; they can be portrayed in any awkward or embarrassing situation and the recording will be convincing.[59] Soon, no real original recording will be required to achieve this.

3.2. Punishment for misstatements—rules of general application

The possibility of prohibiting misstatements, including deliberate lies, and depriving them of the protection of freedom of speech is an issue that touches on several aspects of the legal system, although it can be stated that within the framework of the constitutional protection of freedom of speech, lying—in a general sense and due to its inherently untruthful nature—cannot be prohibited. At the same time, legal action against falsehoods is still possible in certain circumstances, for example, in Hungary in relation to defamation, the denial of the Nazi and Communist genocides, the denial of crimes against humanity, the spreading of scare stories, making false statements in election campaigns, and misleading commercial practices.

3.2.1. The question of the general prohibition of lying

According to the currently used doctrine, lies—in a general sense—cannot be prohibited under the framework of the protection of freedom of speech. Ferenc Deák, the prime minister of the first independent Government of Hungary, would certainly

57 Wu, 2017; Hindman, 2019.
58 Park, 2018, p. 7.
59 Chesney and Citron, 2019.

be disappointed to hear this. Upon the codification of the first Hungarian Press Act in the spring of 1848,[60] he asserted that "If it were up to me, there would be only a single paragraph in the Press Act: 'Lying is forbidden'" (although, who knows, maybe this quotation is itself fake news; after all, everyone quotes the pithy saying without citing the original source). Today's press and media acts impose numerous obligations on journalists, save for the general prohibition of lying. However, this does not preclude broad restrictions or post-publication sanctions being imposed for false statements of facts.

3.2.2. Protection of reputation

One of the most important areas of the legal protection of human personality is the law of libel, the defamation law that serves the protection of reputation and honor and is intended to prevent unfavorable and unlawful changes to an individual's image and external social perception. By means of these rules, the legal system aims to prevent any opinion published in the public sphere from damaging or even ruining an individual's image without proper grounds, primarily through false statements. On this question, individual states' approaches are remarkably diverse, but the common point of departure in Western legal systems is the strong protection of debates on public affairs; as such, the protection of the personality rights of public figures is forced into the background when compared to the protection of the freedom of speech.

The boundaries of the protection of the personality rights of public figures are primarily shaped by the decisions of the courts and the CC. As it relates to the constitutional protection of honor and reputation, the idea that statements relating to public affairs and damaging public figures' reputation may legitimately claim some special protection only gained ground in Hungary after 1989. Initially, the limited protection of public figures' personality rights was not based on statutory provisions. The point of departure on this issue was CC decision 36/1994. (VI. 24.) AB, which laid down the principles that serve as its foundation. According to the CC's position, the possibility of publicly criticizing the activity of bodies and persons fulfilling state and local government tasks is an outstanding constitutional interest, as is ensuring that citizens may participate in political and social processes without uncertainty, compromise, and fear. As such, while the constitutionality of protecting the honor and reputation of such public figures by means of criminal law may not be excluded, the freedom of opinion—in comparison to that of private persons—may be limited to a rather narrow extent, only in order to protect those exercising state powers.

This test for establishing the perpetrator's liability for deliberate lies or in the event of negligence is rather similar—but not identical—to 'the *New York Times* rule' developed in *New York Times v. Sullivan*. The codification of the Civil Code in 2013 is a milestone in the context of the possibility of limiting the protection of public figures' reputation and honor under civil law, as a result of which the legislature

60 Act XVIII of 1848.

adopted statutory provisions to settle the issue of the protection of public figures' personality rights (Article 2:44). The Civil Code in its current form provides a broader framework for the discussion of public affairs and also defines its limits: The publication of opinions that offend human dignity cannot be considered to be speech protected by the freedom of expression, regardless of the status of the person concerned or the public nature of the issue under discussion. CC decision 7/2014. (III. 7.) AB, clarifying the test for protection of reputation, stipulated for statements of fact that "demonstrably false facts in themselves are not protected by the Constitution" (Paragraph [49]), thereby implying that, in certain cases, even false statements of facts can receive protection under the freedom of opinion. The decision also establishes that:

> Even for those facts having no constitutional value which later turn out to be false, it is justified to take into account the interest of ensuring as free conditions for the discussion of public affairs as possible when determining the extent of imputability (attribution of liability) and the possible penalties in the course of the legal proceedings. (Paragraph [50])

3.2.3. Genocide denial

According to the Council Framework Decision on combating racism and xenophobia in the EU member states, a universal prohibition shall be applied to the denial of crimes against humanity, war crimes, and genocides.[61] Most member states have a law prohibiting the denial of the crimes against humanity committed by the national socialists, questioning them, or downplaying their significance.[62] According to the currently effective Hungarian regulation, if a person "in front of a large audience, denies, questions, belittles or seeks to justify the genocide and other crimes against humanity committed by the national socialist and communist regimes [he] is guilty of a felony" (Article 333 of the Criminal Code). The CC did not declare this provision of the Criminal Code to be unconstitutional. According to the statement of reasons:

> Denying the sins of Nazism and Communism shall be considered as abuse of freedom of expression, which severely injures not only the dignity of the community of victims but the dignity of citizens committed to democratic values and identifying with or sympathising with the victims.[63]

In addition to protecting human dignity (individual rights), the CC also considered the protection of public peace to be important, at the same time avoiding the controversial question of the degree of threat to public peace that may justify the

61 Council Framework Decision 2008/913/JHA.

62 See the French Gayssot Act (13 July 1990, amending the Press Act of 1881, by adding a new Section 24), and the German Strafgesetzbuch, s. 130(3).

63 Decision 16/2013 (VI. 20.) AB, para. 50.

restriction of a fundamental right and using the protection of public peace only as a secondary argument.

3.2.4. Scaremongering

Hungarian criminal law has sanctioned the dissemination of scaremongering (originally 'frightening rumors')[64] since the end of the nineteenth century. Based on the currently effective legislation, this offense is committed by anyone who, in a place where it constitutes a public danger, states or reports an untrue fact or distorts a true fact in a manner likely to cause confusion or disquiet among a large group of people in the place of public danger. In addition, since the spring of 2020, as one of the measures introduced relating to protection against the novel coronavirus, a new provision has been added to the law: "It is also considered spreading scare-stories if someone, during a special legal regime, states or spreads false facts or distorts true facts in a way that is likely to hinder or frustrate the effectiveness of the defence" (Articles 337(1)-(2) of the Criminal Code).

The CC, in its decision 15/2020. (VII. 8.) AB, aimed to assess the constitutionality of the regulation and concluded that the provision meets the constitutional requirements. The prohibition pertains only to a specific category of statements of facts. The scope of (untrue) information that could impede the effectiveness of the defense against the pandemic is relatively narrow, at least much narrower than all the published statements of fact in relation to the threat justifying the introduction of a special legal regime. The prohibited action must be objectively capable of hindering or frustrating the effectiveness of the defense, whether undertaken by the government or by other public, municipal, or even private actors acting in concert (Paragraphs [53], [60], [63]). To assist the application of the law, the decision lays down a constitutional requirement for a range of statements of facts and strengthens the protection of freedom of opinion. This is the case where the truth of a statement of fact contained in a communication cannot be established at the time of communication but subsequently proves to be false:

> The statement of fact can only be punished if it is a statement of a fact that the offender must have known to be false at the time when the act was committed or that he himself distorted and that is capable of hindering or frustrating the defence during the special legal regime. (Operative part, Section 1)

3.2.5. Election procedures and political advertisements

Numerous specific rules apply to statements made during election campaigns. These can have a twofold purpose. On the one hand, they powerfully protect communication during a campaign: Political speech is the most strictly protected inner

64 See Act XL of 1879, art. 40.

core of freedom of speech, and what is said in a campaign is intimately related to the functioning of democracy and democratic procedures. On the other hand, these procedures must also be protected, so that a candidate, community, or party does not distort the democratic decision-making process and ultimately harm the democratic order. It is no coincidence that the fake news problem becomes most visible during election campaigns (e.g., the US presidential elections, the 2019 European Parliament elections, etc.).

Many European countries have laws in place that limit the publication of political advertisements in terms of their quantity, the equal distribution of media space, clients that can commission them, or the amount of money that can be spent on them. Their main purpose is to ensure a level playing field to the detriment of parties and candidates with greater financial resources and for the benefit of less privileged parties. In this respect, the Hungarian legislation is mixed. On the one hand, the possibility of publishing political advertisements in the media (television, radio) is heavily regulated,[65] while on the other hand, if the request for publication of the communication complies with the relevant legal provisions, media service providers are obliged to publish it without any discretion, i.e., they are not liable for the content of the political advertisements they publish.

Article 2(1)e of Act XXXVI of 2013 (the Election Procedures Act) requires all parties involved in the elections to exercise "their rights in a bona fide manner and with the proper intent," and this requirement also includes the prohibition of disseminating false statements. However, the messages communicated during election campaigns belong to the most protected sphere of expression of opinion; therefore, if the statements made during election campaigns concern public figures and relate to their political activities, program, or credibility and suitability, it may be assumed that voters will deem these statements to be opinions, even if the statements were formulated in the indicative mood. With due consideration of this, the consideration of individual statements made during an election campaign clearly goes beyond an examination of the elements of the statement by applying the provability test and requires the evaluation of all the conditions relating to the case.[66]

Furthermore, according to the case law, during an election campaign "freedom of expression must typically be interpreted and judged in the context of the interplay

65 Hence, for instance, the publication may take place only during the election campaign period or in connection with an ordered referendum (Media Act, art. 32(3)) and free of charge (Fundamental Law, art. IX(3) and Act XXXVI of 2013 on the Election Procedures (the Election Procedures Act), art. 147(3)); the media service provider is not allowed to express its opinion or provide any evaluative explanation to the political advertisement (Election Procedures Act, art. 147(2)); the person/entity commissioning the respective political advertisement must be clearly defined (Media Act, art. 32(4)); the public service media provider is obliged to publish political advertisements for a certain duration and under certain conditions (Election Procedures Act, arts. 147/A-E); the national commercial media service provider is obliged to indicate until a pre-defined time if it intends to participate in the campaign, also indicating the duration intended for publication (Election Procedures Act, art. 147/F).

66 Decision 3107/2018 (IV. 9.) AB.

between public figures" and competing candidates "may express themselves openly and even in an unvarnished manner." If it becomes clear from the context of the statement to be assessed that the speaker did not intend to make a statement of fact but rather to express an opinion, the veracity of the communication should not be assessed, and consequently, publication cannot be restricted in this respect either. Therefore, the freedom of expression is "given increased protection in relation to value judgements which surface in a collision between opinions on public affairs, even if they are perhaps exaggerated and heightened."[67]

3.2.6. Advertising regulation and consumer protection

Advertisements do not exclusively serve the advertiser's interests but may equally serve the interests of the recipients too, hence genuine and fair commercial communication enjoys protection. False or misleading statements published with a commercial purpose may not be afforded constitutional protection, however. This also follows from the European consumer protection rules.[68] In Hungary, both legal regulations implementing the EU rules and the industry-specific systems of self-regulation explicitly prohibit commercial practices communicating false information or communicating true information in a deceptive manner to consumers.[69] In order to prevent unfair competition, the Hungarian Competition Act prohibits communications that damage or jeopardize market participants' reputation and credibility by stating or disseminating untrue facts or by misrepresenting a fact.[70]

3.3. Correcting falsehood—media regulation tools

Media regulation applies to television, radio, on-demand audiovisual services, and radio media services, some of which may also be distributed on the Internet, although social media platforms are not covered by the regulation. With that said, it is worth reviewing the means by which media regulation tries to fight against false statements, as some of these may also serve as a model, or at least an inspiration, for regulating the platforms.

3.3.1. The principle of media pluralism

The regulation of television, radio, on-demand audiovisual services, and radio media services seeks to correct—by indirect means—public communication that may be distorted by the publication of falsehoods. In line with the theoretical requirement for media pluralism, the entire media market must provide collectively

67 Decision 5/2015. (II. 25.) AB.
68 Directive 2005/29/EC, arts. 6, 7.
69 See: Act XLVII of 2008, arts. 6, 7 and Annex; Hungarian Code of Advertising Ethics, art. 10.
70 Act LVII of 1996, art. 3.

for the diversity of opinions and available content, and establish a balance among them.[71] This condition primarily imposes tasks on the state in respect of the regulation of the media market and these regulations mainly concern traditional broadcasting of television and radio services.

Article IX(2) of the Fundamental Law states that "Hungary recognises and protects the freedom and diversity of the press, and ensures the conditions for free information necessary for the development of a democratic public opinion." Media regulation therefore not only establishes the importance of promoting pluralism in the media market and preventing information monopolies as a fundamental principle, but also sets a requirement to prevent unjustified restrictions of competition in the market (Article 4 of the Media Act). This rule of the Fundamental Law concerning the protection of the freedom and diversity of media also includes the obligation to fight against the possible establishment of information monopolies.

Media regulations seek to guarantee the pluralism of the media market and the prevention of the emergence of a dominant position by various means. On the one hand, in order to prevent the emergence of information monopolies, it imposes media market ownership restrictions on media outlets (media service providers) with an annual average audience share above a predefined threshold. On the other hand, it imposes extra obligations on media service providers that reach a certain audience share that can guarantee the maintenance of the diversity of the media market, as well as the promotion of the right to objective, pluralistic information and access to information (Article 67 of the Media Act). The other requirement aimed at preventing concentration of ownership concerns the market for radio media services. This requirement stipulates that media with a national, regional, or local reception area may only be owned by certain undertakings subject to certain limits: A media service provider may hold a maximum of one national, or two regional and four local, or twelve local media service provision rights (media licenses) at the same time (Article 71 of the Media Act).

3.3.2. The right of reply

Based on the right of reply, the legislator does not grant access to a media service provider's content due to some external condition, but due to the service provider's (previously) published content. Article 28 of the AVMS Directive makes it mandatory for EU member states to have legal regulations in place in respect of television broadcasting that ensure adequate legal remedy for those whose personality rights are infringed by the publication of false statements. Such regulations are known Europewide, typically imposing obligations on printed and online press alike.[72] The Hungarian legislator has complied with the relevant Recommendation of the European

71 See Komorek, 2012.
72 Youm, 2008; Richter, 2018–2019, pp. 14–19; Koltay, 2013.

Parliament[73] under the Press Freedom Act, and as such, the right of reply obligation of media content providers falling under the scope of media regulations is also included in codified law.

In Hungary, the institution of the right of reply guarantees that if a false fact is stated or disseminated or a true fact is represented as false concerning a person, the person concerned has the right to demand the publication of a corrective statement (Article 12 of the Press Freedom Act). The right of reply can thus only be enforced in relation to statements of fact (hence, it is not granted for opinions that may be considered offensive), and this serves two purposes simultaneously: both protecting the person affected by the media content and satisfying the public's need to access information.

Under the legislation, any statement (whether it be untrue or true but misrepresented) published in any media content can be used as grounds to enforce the right of reply. According to Point 7 of Article 1 of the Media Act, media content comprises content published in media services and press products, meaning that this notion includes linear and on-demand media services, as well as print and online press products, including all content published in these. Based on the regulations, the right of reply (publication of corrective statements) can only be requested in respect of content that falls under the scope of the media regulation. Thus, for instance, since reader comments on articles published in online press products do not constitute edited content in principle—without prior moderation, that is, without a conscious editorial decision to publish them—and are therefore not part of the press product itself, hence they are not considered media content, there is no right of reply against them either. Similarly, there is no right of reply against newspapers that do not meet the definition of a 'press product' as defined in the Press Freedom Act, particularly those that are not provided as a 'business service' (such as blogs, university student newspapers, association newsletters, political party websites, etc.).

3.3.3. The obligation of impartial news coverage

Among the rules that support media pluralism is the requirement of impartial news coverage, which requires that information programs report on public affairs in an unbiased and balanced manner. Regulation may apply to both television and radio broadcasters, and it is implemented in several states in Europe.[74] The obligation pertaining to the diversity of the press, as stipulated under Article IX(2) of the Fundamental Law, is achieved partly through the requirement of impartial coverage. The content of this obligation is specified jointly by Article 13 of the Press Freedom Act and Articles 12 (1)-(2) of the Media Act. The requirement of impartial coverage is laid down in the Press Freedom Act in relation to information and news programs

73 Recommendation 2006/952/EC.

74 See, for instance, the German (Rundfunkstaatsvertrag, ss. 25-34) and the British regulations (Communications Act 2003, s. 319(2)c-d, s. 319(8) and s. 320; Broadcasting Code, s. 5).

published by linear media services providing information services. Under the provisions of the Act, such programs should provide balanced coverage of local, country-level, national and European issues that may be of interest to the general public and on any events and debated issues bearing relevance to the citizens of Hungary and the members of the Hungarian nation.

The requirement of impartial news coverage—which stems from the recognition of the media's public interest duties– has been an integral part of Hungarian media regulation since its inception dating back to 1996. Television and radio are thus required to provide balanced coverage of events of public interest and controversial issues (Article 13 of the Press Freedom Act). Accordingly, based on the rule, information on and coverage of community affairs must represent the opposing views. The relevant positions on a given issue have to be collected and presented to the public, enabling them to reach a well-founded decision on the issue, thereby also furthering the ideal of democracy. It should be noted that according to the relevant interpretation of the CC, the requirement of impartial coverage is not contrary to the fundamental right of freedom of the press; however, it is a constitutional requirement that the balanced nature of coverage must be examined within individual programs or in the entirety of the programs as a whole, depending on the nature of the program (see: CC decision 1/2007. (I. 18.) AB). It is important to emphasize that certain communications containing an untrue element cannot in themselves lead to the establishment of a breach of this legal provision, only if they also result in a breach of the requirement of impartial coverage, as defined above.

4. Summary

The legal relationship between social media platforms and their users (which is not affected by the constitutional doctrines of free speech) is also governed by law through the contract concluded by and between the parties. However, it does not seem possible to enforce the principles and doctrines of free speech in the online world with the same fervor as can be done offline. With the rise of the Internet, the right to the freedom of speech seems to have entered a new phase of development, the future consequences of which we can only guess at today.

Government decision makers and shapers of public policy need to adopt a systemic approach that considers the distinctive features of gatekeepers' activities, keeps track of the changes in them, provides an accurate definition of what gatekeepers are expected to do and what they might expect from the law, and accurately lays down gatekeepers' duties and the scope of their liability. Gatekeepers' impact on public communication and the strengthening of private regulation necessitate the use of new, creative, and innovative regulatory methods and institutions, the invention

of new ways of establishing rules, and a degree of cooperation between public and private actors that is unprecedented in this field.

As far as the problem of fake news is concerned, the current doctrine of freedom of speech as applied in Europe does not exclude the prohibition of the publication of falsehoods, hence these cannot enjoy general constitutional protection. False statements of fact can, in certain cases, be restricted. However, their general prohibition is hard to imagine. At the same time, this is a serious and massive problem for public communication and the discussion of public affairs, especially on large online platforms. Any possible regulation is either contrary to the principles of freedom of speech or is likely to be ineffective. For the time being, states seem to accept that they will not be able to regulate the public sphere without the platforms themselves, and they are deliberately handing over to the platforms their former exclusive state function of setting the boundaries of freedom of speech.

Bibliography

BALKIN, J. (2018) 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *University of California Davis Law Review,* 51, pp. 1149–1210.

BARENDT, E. (2005) *Freedom of Speech.* Oxford: Oxford University Press, pp. 155–162.

Beperli a Facebookot a Mi Hazánk elnöke [Online]. Available at: https://mandiner.hu/cikk/20200429_beperli_a_facebookot_a_mi_hazank_elnoke?utm_source=mandiner&utm_medium=link&utm_campaign=mandiner_202104 (Accessed: 14 May 2021)

BERNAL, P. (2018) *The Internet, Warts and All. Free Speech, Privacy and Truth.* Cambridge: Cambridge University Press

BÉNYEI, M. (1994) *Reformkori országgyűlések a sajtószabadságról.* Debrecen: Stúdium

Bizalom, tudatosság, veszélyérzet az interneten [Online]. Available at: https://itki.uni-nke.hu/hirek/2020/06/17/bizalom-tudatossag-veszelyerzet-az-interneten (Accessed: 14 May 2021)

BRAGYOVA, A. (2016) 'Szólásszabadság és sajtószabadság az internetes kommentek példáján' in Fejes, Zs., Török, B. (eds.): *Suum cuique: ünnepi tanulmányok Paczolay Péter 60. születésnapja tiszteletére.* Szeged: Pólay Elemér Alapítvány, pp. 473–490.

CHESNEY, B., CITRON, D. (2019) 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security', *California Law Review,* 107, pp. 1753–1819.

DARNTON, R. (2017) 'The True History of Fake News' [Online]. Available at: https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news (Accessed: 14 May 2021)

Ez már hajtóvadászat: a Google törölte a harmadik Youtube-csatornánkat is [Online]. Available at: https://pestisracok.hu/ez-mar-hajtovadaszat-a-google-torolte-a-harmadik-youtube-csatornankat-is (Accessed: 14 May 2021)

FRADETTE, J. E. (2013–2014) 'Online Terms of Service: A Shield for First Amendment Scrutiny of Government Action', *Notre Dame Law Review,* 89(2), pp. 947–984.

GILBERT, D. (2018) *Why Facebook Censored a Racist Video from Hungary's Government – Then Put it Back* [Online]. Available at: https://www.vice.com/en/article/gy87m4/why-facebook-censored-a-racist-video-from-hungarys-government-then-put-it-back (Accessed: 14 May 2021)

GRAD-GYENGE, A. (2015) ' Commentare necesse est – néhány gondolat az Alkotmánybíróság komment-határozatáról', *Glossa Iuridica*, 1–2, pp. 122–142.

Gyerekpornós részlet miatt törölték a Pesti Srácok YouTube-csatornáját [Online]. Available at: https://hvg.hu/itthon/20200211_gyermekpornografia_pesti_sracok_video_youtube_csatorna (Accessed: 14 May 2021)

HEINS, M. (2014) 'The Brave New World of Social Media Censorship', *Harvard Law Review Forum,* 127, pp. 325–330.

HINDMAN, M. S. (2019) *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy.* Princeton: Princeton University Press

House of Lords (2014) *Social Media and Criminal Offences* [Online]. Available at: https://publications.parliament.uk/pa/ld201415/ldselect/ldcomuni/37/3702.htm (Accessed: 14 May 2021)

KLEIN, T. (2016) 'A tárhelyszolgáltató „omnipotens" felelőssége mint alkotmányjogi problematika: A harmadik személy tartalmáért való szolgáltatói felelősség az interneten' in Koltay, A., Török, B. (eds.) *Sajtószabadság és médiajog a 21. század elején 3.* Budapest: Wolters Kluwer. pp. 347–374.

KOHL, U. (2018) 'Islamophobia, "Gross Offensiveness" and the Internet', *Information & Communication Technology Law,* 27(1), pp. 111–131.

KOLTAY, A. (2013) 'The Right of Reply in a European Comparative Perspective', *Acta Juridica Hungarica – Hungarian Journal of Legal Studies,* 1, pp. 73–89.

KOLTAY, A. (2015) 'Az Alkotmánybíróság határozata az internetes kommentek polgári jogi megítéléséről', *Jogesetek Magyarázata,* 1, pp. 9–22.

KOMOREK, E. (2012) *Media Pluralism and European Law.* New York: Wolters Kluwer

MRSZ Barométer – a válság hatásai a reklámiparban [Online]. Available at: https://mrsz.hu/cmsfiles/3c/dd/MRSZ_Barometer_iparagi_kutatas_2020.szeptember_v09.pdf (Accessed: 14 May 2021)

PARK, M. K. (2018) 'Separating Fact from Fiction: The First Amendment Case for Addressing "Fake News" on Social Media', *Hastings Constitutional Law Quarterly,* 46(1), pp. 1–15.

PÁLFY, D. Á. (2020) *Bepereljük a Facebookot. Megint. Toroczkai László a Mandinernek* [Online]. Available at: https://mandiner.hu/cikk/20201101_bepereljuk_a_facebookot_megint_toroczkai_laszlo_a_mandinernek (Accessed: 14 May 2021)

Remove, Reduce, Inform: New Steps to Manage Problematic Content [Online]. Available at: https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps (Accessed: 14 May 2021)

RICHTER, A. (2018–2019) 'Fake News and Freedom of the Media', *Journal of International Media & Entertainment Law,* 8(2), pp. 1–33.

ROBB, A. (2017) *Anatomy of a Fake News Scandal* [Online]. Available at: https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877 (Accessed: 14 May 2021)

ROBINSON, D. (2017) *Social Networks Face Tougher EU Oversight on Video Content* [Online]. Available at: https://www.ft.com/content/d5746e06-3fd7-11e7-82b6-896b95f30f58 (Accessed: 14 May 2021)

ROWBOTTOM, J. (2012) 'To Rant, Vent and Converse: Protecting Low Level Digital Speech', *Cambridge Law Journal,* 71(2), pp. 355–383.

SCHAUER, F. (1998) 'The Ontology of Censorship' in Post, R. C. (ed.) *Censorship and Silencing: The Practices of Cultural Regulation.* Los Angeles: GRI. pp. 147–168.

SEPSI, T. (2015) 'No comment? Legal liability for online comments', *Fundamentum,* 4, pp. 106–110.

SUBRAMANIAN, S. (2017) *Inside the Macedonian Fake-News Complex* [Online]. Available at: https://www.wired.com/2017/02/veles-macedonia-fake-news (Accessed: 14 May 2021)

SZIGETI, T., SIMON, É. (2016) 'A hozzászólás szabadsága: a közvetítő szolgáltatói felelősség aktuális kérdéseiről', *Fundamentum,* 2-4, pp. 113–124.

Toroczkai törlése után megregulázná a Facebookot a Mi Hazánk [Online]. Available at: https://24.hu/kozelet/2019/07/01/mi-hazank-mozgalom-facebook-vizsgalobizottsag (Accessed: 14 May 2021)

Törölte a Mi Hazánk oldalát a Facebook [Online]. Available at: https://hvg.hu/itthon/20201027_torolte_a_mi_hazank_oldalat_a_facebook (Accessed: 14 May 2021)

VARGA, J. (2021) Facebook-post [Online]. Available at: https://www.facebook.com/VargaJuditMinisterofJustice/posts/4049403765078549 (Accessed: 14 May 2021)

YOUM, K. Ho (2008) 'The Right of Reply and Freedom of the Press: An International and Comparative Perspective', *George Washington Law Review,* 76(4), pp. 1017–1064.

WU, T. (2017) *Attention Merchants. From the Daily Newspaper to Social Media. How our Time and Attention is Harvested and Sold.* London: Atlantic Books

Hungarian legal sources

Fundamental Law of Hungary
1/2007. (I. 18.) AB (Constitutional Court)
5/2015. (II. 25.) AB (Constitutional Court)
15/2020. (VII. 8.) AB (Constitutional Court)
16/2013 (VI. 20.) AB (Constitutional Court)
19/2014. (V. 30.) AB (Constitutional Court)
20/1997. (III. 19.) AB (Constitutional Court)
36/1994. (VI. 24.) AB (Constitutional Court)
3107/2018. (IV. 9.) AB (Constitutional Court)
Act CIV of 2010 on the Freedom of the Press and the Fundamental Rules of Media Content
Act CVIII of 2001 on certain issues of electronic commerce services and information society services
Act CLXXXV of 2010 on Media Services and Mass Media
Act II of 1986 on the Press
Act LVII of 1996 on the Prohibition of Unfair Market Practices and Restriction of Competition
Act XVIII of 1848 on the Press
Act XL of 1879 on Petty Offences
Act XLVII of 2008 on the Prohibition of Unfair Commercial Practices against Consumers
Act XXXVI of 2013 on the Election Procedures
Act V of 2013 on the Civil Code
BDT2013.2904 (Pécs Court of Appeal)
BH2016.330 (Curia)
BDT2017.3675 (Pécs Court of Appeal)
BDT2019.3987 (Szeged Regional Court of Appeal)
Code of Advertising Ethics

EU legal sources

C-18/18 Eva Glawischnig-Piesczek v. Facebook Ireland Ltd., judgment of 3 October 2019
Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online [C(2018) 1177 final]
Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expression of racism and xenophobia by means of criminal law
Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, particularly electronic commerce, in the Internal Market
Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market (Unfair Commercial Practices Directive)

Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (AVMS Directive), amended by Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018

Recommendation of the European Parliament and of the Council of 20 December 2006 on the protection of minors and human dignity and on the right of reply in relation to the competitiveness of the European audiovisual and on-line information services industry (2006/952/EC)

Other legal sources

Broadcasting Code (the United Kingdom)

Communications Act 2003 (the United Kingdom)

Delfi v. Estonia no. 64569/09, judgment of 16 June 2015 [Grand Chamber]

Loi du 29 juillet 1881 sur la liberté de la presse (France)

Loi no 90-615 du 13 juillet 1990 tendant à réprimer tout acte raciste, antisémite ou xénophobe (Gayssot Act, France)

Magyar Tartalomszolgáltatók Egyesülete and Index.hu. Zrt v. Hungary no. 22947/13, judgment of 16 February 2016

New York Times Co. v. Sullivan 376 U.S. 254 (1964)

Staatsvertrag für Rundfunk und Telemedien (Rundfunkstaatsvertrag, Germany)

Chapter IV

The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism in Serbia

Sanja Savčić

1. Introduction

Communication is a natural human need. It is based on the inner and outer elements. The first concerns cognition and the thinking process, and the second refers to communicating information and formed opinions to others. Viewed in this way, communication is a continuous exchange of opinions, facts, and knowledge. This exchange is the basis for connecting but also separating individuals, groups, and even entire social communities. Essentially, such communication is the basis of civilizational achievements and social and political changes.

At the beginning of the 21st century, social networks expanded in the communication domain. Among them, the most popular are Facebook, MySpace, Twitter, and TikTok. These and similar services have become part of a global trend within a very short time, even though most were initially established as university networks. At present, there are almost no Internet users who do not use at least one social network. The development of activities on social networks can be compared with the development of technologies with both high-processing power and the memory capacity for storing the vast amount of data exchanged via the Internet. New information technologies have enabled new ways of applying and communicating information. The development of social media has enabled the rapid dissemination of

Sanja Savčić (2021) The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism in Serbia. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 111–140. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_4

information and interaction between users and the community. Apart from these digital advantages and besides the fact that the Internet has made daily life much more comfortable, the digital environment and the online environment in particular have put lawmakers in a very unfavorable position from the outset. The new environment has exposed traditional legal concepts' weaknesses and consequently raised a lot of questions. One pertains to freedom of expression and its exercise in the digital (i.e., the online) environment, especially through social networks.

The term 'social network' is most often defined as a social structure composed of individuals (or organizations) that are connected by one or more specific types of interdependence, such as values, visions, ideas, financial interests, friendship, kinship, common interests, financial exchange, sexual relations, or relations of trust, knowledge, or prestige. These virtual communities, which are available to all Internet users, enable fast communication and the exchange of various content, such as photos, videos, news, and events. Today, social networks are used by hundreds of millions of people worldwide. That has led to the creation of a new type of society, the so-called information society.

Bearing this in mind, there is no doubt that social networking's impact on modern life is extensive, but the question is to what extent this virtual world changes perceptions of modern society's recognized values. From the legal point of view, the most impacted area is the one regarding human rights, specifically the freedom of expression.

1.1. Freedom of expression

Freedom of expression is a first generation political and civil right. It includes all forms of expression: verbal (freedom of speech, as such), printed (freedom of the press), and artistic.

Freedom of expression was proclaimed in 1789 in the French Declaration of the Rights of Man and the Citizen as an inalienable right. Article 11 of this Declaration reads:

> The free expression of ideas and opinions is one of the most precious human rights. Accordingly, every citizen can speak, write and print freely, but he will also be responsible for the abuse of this right, in the manner prescribed by law.

Freedom of expression is also protected by all international instruments dealing with the protection of human rights, and for the first time, it has been proclaimed via an international act, the Universal Declaration of Human Rights,[1] which states that:

> Everyone has the right to freedom of opinion and expression, including the right not to be disturbed by his opinion, as well as the right to seek, receive and impart information and ideas by any means and regardless of frontiers.

1 Adopted by the General Assembly of the United Nations, on 10 December 1948.

Furthermore, this freedom, in addition to other civil and political rights, was proclaimed by the International Covenant on Civil and Political Rights, which was adopted by the General Assembly of the United Nations on December 16, 1966.[2] Art. 19 states that:

> Everyone has the right to freedom of expression. This right includes freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, whether oral, written or printed, in the form of art or other media, or any other means of choice. The use of (these) freedoms (…) entails special duties and responsibilities. It may, therefore, be subject to certain restrictions which must nevertheless be expressly provided for by law and which are necessary for: a) respect for the rights and reputation of other persons; b) protection of national security or public order, or public health or morals.

Apart from the fact that freedom of expression is a right in itself, it is an indivisible element of other rights, such as political rights, rights of education, intellectual property rights, etc., or the means of their exercise. From that point of view, the passive holding of the state is not enough to render this right effective. It is equally important to provide complete protection for this right and enable information exchange.

1.2. Freedom of expression and digital platforms – Introductory remarks

Social media, including social networks and the Internet in general, are a convenient arena for this exchange. In terms of law, it is the realization of two mentioned rights: to express thoughts and to be informed.[3] Keeping the nature of the Internet in mind, it is not incorrect to say that social networks comprise a virtual arena of freedoms, intensified by the fact that the laws, rules, and regulations that often govern us in real life are unclear, inapplicable, or absent.

Even though social networks have emerged as technical platforms and digital spaces for their users, they have become much more than that, and in so doing, are much closer to the term 'media.' Their improved function has been recognized through attempts to self-edit content. This corresponds with users' active role. In addition to providing only an information structure for information exchange, their visibility and distribution, alongside numerous abuses, have led to the question of how to edit. The need to edit something leads social networks into the world of media. Content on social networks, which is created by the users themselves, has become public space and exerts an influence on public opinion.

However, unlike the media, content on social networks is created by users, that is, ordinary citizens, with no real sensibility for the seriousness of and consequences

2 Etinski, 2004, p. 276.
3 Mijović, 2020, pp. 1026–1027.

associated with the opinions they express or the information they share. Their contributions are simply raw. Following the increasing incidence of mistreatment, the need to introduce conditions of use that could curb hate speech, insults, and threats has arisen. This is obviously the case when it comes to the prevention of abuse and harassment and the safety of children, who are especially vulnerable. Therefore, the question is: Who should edit the content distributed on social networks, i.e., who is supposed to be tasked with controlling it?

As regards social networks, self-regulation is offered as well, but it is not enough. There are at least three reasons for this remark. Firstly, it is impossible to ignore the fact that the companies that own and manage the infrastructure and enact regulations do not have influence over the users who create the content independently. Even if this was not the case, companies are naturally disinterested in setting unpopular rules that could repel users from social networks. Secondly, public opinion on social networks is of public interest and is therefore important to the authorities. Finally, the Internet erases state boundaries, while companies are subject to a particular country's legal system. That leads to the conclusion that countries cannot protect public interest, since companies do not have to be under state jurisdiction. Hence, it is crucial for authorities to intervene via state regulation.

This chapter will observe to what extent Serbian law can respond to the challenges of modern virtual society and preserve traditional human rights. For this purpose, the legal framework for freedom of expression, as well as its limitations, will be introduced. Afterward, methods of monitoring Internet content pursuant to various legal acts will be offered.

2. Legal aspects of content censorship on social networks in Serbia

2.1. Introductory remarks

In modern society, information is the most dominant commodity on the market, as well as in politics, national safety, culture, and science. However, information itself is neutral. From that perspective, the information must be available in order to be used. The provision and exchange of information form the core of communication. However, freedom of expression is the basis for enabling information exchange.

Freedom of expression includes all forms of expression: verbal (freedom of speech, as such), printed (freedom of the press), and artistic. In that sense, each modern and democratic country has an obligation to refrain from actions that could obstruct the exercise of this freedom through censorship or other restrictions. Apart from the fact that freedom of expression is a right in itself, it is an indivisible element of other rights, such as political rights, rights of education, intellectual property rights,

etc., or the means of their exercise. From that point of view, the passive holding of a country is insufficient to render this right effective. It is equally important to provide complete protection for this right and enable information exchange.

2.2. Freedom of expression in Serbia

Art. 46 of the Constitution of the Republic of Serbia states that the freedom of thought and expression shall be guaranteed, as well as the freedom to seek, receive, and impart information and ideas through speech, writing, art, or in some other manner. Freedom of expression may be restricted by the law if necessary to protect the rights and reputation of others, to uphold the authority and objectivity of the court, and to protect public health, the morals of a democratic society, and the national security of the Republic of Serbia.[4] On the other hand, freedom of thought, as an internal psychological process, cannot be limited by law, since that process is out of cognition and hence outside the authorities' control.[5]

Further, in Art. 50, titled *Freedom of the media*, the Constitution of the Republic of Serbia states that everyone shall have the freedom to establish newspapers and other forms of public information without prior permission and in a manner defined by the law that regulates the particular matter. The same applies to the establishment of television and radio stations.

Censorship shall not be applied in the Republic of Serbia (Art. 50(3)), but prevention of the spreading of information through means of public informing is not excluded. That may happen only when a competent court finds it necessary. In accordance with the principles of democratic society, a justified aim is to prevent inciting the violent overthrow of the system established by the Constitution or to prevent the advocacy of racial, ethnic, or religious hatred fueling discrimination, violent hostility, the exercise of the right to correct false, incomplete, or inaccurately imparted information resulting in the violation of any person's rights or interests, and the right to react to communicated information (Art. 50(4)).

Shaping an opinion that is supposed to be published depends on the facts the person has obtained. From that perspective, freedom of thought and expression as well as freedom of the media are as important as the right to be informed. Hence, the Constitution of the Republic of Serbia contains an explicit provision that guarantees everyone the right to be informed accurately, fully, and in a timely manner about matters of public importance. The media shall have the obligation to respect that right (Art. 51(1)). When it comes to the information kept by state bodies and organizations with delegated public powers, the right to access information shall be exercised in accordance with the specific law (51(2)).

4 Ustav Republike Srbije (Constitution of the Republic of Serbia), Official Gazette of the Republic of Serbia, No. 98/2006, dated on 8 November 2006. English version of the Constitution is available at: https://www.propisi.pravno-informacioni-sistem.rs/content.php?id=800.

5 Orlović, 2019, p. 90.

Taking the provisions of the highest legal act into account, it is clear that freedom of expression, coupled with the correlated right to information, is not an unrestricted right. In general terms, those rights could be restricted by law, if the Constitution permits such restriction, for the purpose allowed by the Constitution, and to the extent needed to meet the constitutional purpose of restriction in democratic society without encroaching upon the substance of the relevant guaranteed right. The attained level of human and minority rights may not be lowered.

As relevant provisions provide, freedom of expression, in a broad sense, shall not endanger rights that promote higher values, that is, the rights and reputation of others, as well as the protection of national security, public order, public health, and morals. On the relevance of those rights regarding the individual's dignity, reputation, and honor, it can be testified that freedom of thought and expression is among the rights that cannot be derogated during war or state of emergency (Art. 202(4)).

2.3. Restrictions on freedom of expression

The Constitution of the Republic of Serbia protects freedom of expression as a fundamental right, but it also constitutes a framework for its restriction. In that sense, freedom of expression could be limited by others' rights and reputation and the obligation to uphold the authority and objectivity of the court and protect public health, the morals of democratic society, and the national security of the Republic of Serbia (Art. 46(2)).

This general provision, in fact, has the same meaning as Art. 10 of the European Convention of Human Rights. In that sense, the general provision restricting the freedom of expression has two mutual functions. One is the state's duty to enable any citizen to express their opinion without fear of being punished for it or being put in a different position compared to the unlike-minded. It also includes the accessibility of someone's expression. The other one is the state's duty as well, but it refers to the duty to protect the rights of others.

In order to fulfil both aims, the provisions of several laws prescribe certain content that could result in the violation or jeopardization of others' rights. These are not *numerus clausus* cases. In that sense, if the content is not explicitly recognized as inappropriate, this does not mean it is allowed. In fact, the court that will decide on eventual disputes determines whether the expressed content violates someone's rights or which one of the confronted rights requires protection in a specific situation.

According to a survey on rights violations on the Internet,[6] the most frequent violations are in relation to threatening content, compromising security, insults, and

6 The research was conducted by the Share Foundation through continuous monitoring of violations in online communication. Results in different categories of monitoring are available at https://monitoring.labs.rs/.

unfounded accusations.[7] The affected parties are mostly citizens. However, the reverse is also true, as citizens comprise the majority of attackers. Other analyses consider regional rights violations in the digital environment. The results show that hate speech and discrimination are involved in almost half of the reports of Internet-based digital violations.[8]

Various actions and content could represent the violation and jeopardization of human rights, national security, and the morals of democratic society, but at the same time, this content represents the line between freedom of expression and expression that is punishable by law.

For the purpose of understanding the restriction of the freedom of expression, violations that are prevalent among unlawful actions committed on the Internet will be explained.

2.3.1. Threatening content and compromised security

Content that contains threats to attack another person's life or body could endanger that person's safety. In order for Internet content to be considered dangerous, it must cause fear in the person to whom it is directed, but it must also be directly stated and achievable. In each case, a competent body (usually a court) must determine whether these conditions are met. It is not always easy to correctly determine whether something is an expression of thoughts that are appropriate to jeopardize someone's right (because of a lack of seriousness). Moreover, it is sometimes not obvious which expression among similar ones has the potential to endanger someone's safety, which one causes fear, and finally, whether the attacked person is too sensible for the content. All relevant issues should be evaluated according to the circumstances of the particular case.

7 There are seven different categories of digital right and freedom breaches and several sub-categories under each that have been under monitoring: 1) information security breaches, including making content unavailable through technical means, destruction, and the theft of data and programs, computer fraud, unauthorized access – unauthorized alterations and insertions of content, disabling control over an online account or content; 2) information privacy and personal data breaches (publishing information about private life, illegal interception of electronic communications, breaches of citizens' personal data, illegal personal data processing, breaches of information privacy in the workplace, other breaches of information privacy); 3) pressures because of expression and activities on the Internet (publishing falsehoods and unverified information with the intention of damaging another's reputation, insults and unfounded accusations, threatening content and endangering of security, hate speech and discrimination, freedom of expression on the Internet and the workplace, pressures because of publishing information); 4) manipulation and propaganda in the digital environment (creating fake accounts and paid promotion of false content, content manipulation and organized reporting on social media; changes to or removal of content that is in the public interest; placement of commercial content as news, other manipulation in the digital environment); 5) holding intermediaries liable (pressures because of user-generated content); 6) blocking and filtering of content (blocking/filtering on the network level, algorithmic blocking or suspension of content); 7) other breaches of digital rights and freedoms not included in the categories previously defined. Available at: https://monitoring.labs.rs/.

8 Data available at: https://monitoring.bird.tools/.

To illustrate, the Supreme Court of Serbia, in a decision during a criminal procedure, stated that the defendant, as one of the members of a group advocating violence against individuals belonging to the gay population or organizing a public gathering, sent a threat, the meaning of which was the deprivation of life or at least bodily injury of the participants in an event. The threat was sent through an electronic medium intended for the general public. That means that the threat was available to everyone, that is, a wide range of people, including the mentioned persons to whom the threat was addressed in a certain place. Bearing this in mind, the threat, given the specified circumstances under which it was addressed and could have been known, could objectively create anxiety, insecurity, and fear regarding the personal safety of the members of that circle of persons, who, as passive subjects and possible victims, are indicated in the indictment, in a manner that sufficiently identifies them as members of the target group to which the threat is addressed.[9]

Notwithstanding the outcome of the above case, the court indicates which elements of any statement made on the Internet, on a social network to be precise, are relevant to making the decision.

2.3.2. Insults and unfounded accusations

Expressions on the Internet that violate others' reputation and honor shall be considered unlawful. However, expressions with a negative connotation are not themselves enough to jeopardize someone's right to dignity, reputation, and honor. Devaluing the attacked person is an important element of violation. For that reason, the person toward whom the insult is directed has to be determined. In addition, the statement ought to be insulting according to objective criteria (such as the behavior of the society or group, common opinion on the quality of the statement, and the moral values of the particular group or society), not just the insulted person's perception.[10]

Insults occurring in the media or in the public forum are a qualified crime, and hence stricter punishment is handed down.[11]

2.3.3. Hate speech and discrimination

The Republic of Serbia's Constitution and laws prohibit discrimination. Moreover, it seems that much attention is paid to the violation of others' rights, especially the right of equity. On one side, the reason for this treatment is the fact that equity is a

9 Judgement of the Supreme Court of Serbia, Kzz 59/11, dated 31 August 2011.

10 In Lepojić vs. Serbia, and Filipović vs. Serbia, the European Court of Human Rights stated that authorities have a duty to suffer stronger criticisms and insults, and they need to be much more tolerant than ordinary citizens. This statement was contrary to the opinion of Serbian courts, which claimed that the honor of the authorities is more important than the same personal right of others. Kršikapa, 2008, p. 3.

11 Art. 170, par. 2 Criminal Code of the Republic of Serbia, Official Gazette of the Republic of Serbia, No. 85/05, 72/09, 111/09, 121/12, 104/13, 108/14, 94/16, 35/19. More on this: Petrašinović, 2020, pp. 469–481.

precondition of the exercise of other human rights.[12] There is no value in protecting freedom of religion, for example, if religions are not treated equally by the competent law or in practice. The other reason for this legally developed restriction of freedom of expression is the divergence of behavior or content that could result in discrimination.

According to the Law on Prohibition of Discrimination of the Republic of Serbia,[13] discrimination is any unjustified differentiation, inequitable conduct, or misconduct in relation to persons or groups or members of their families or persons close to them, directly or indirectly, if this behavior is based on race, color, ancestry, citizenship, nationality or ethnic origin, language, religious or political beliefs, gender, gender identity, sexual orientation, property status, birth, genetic characteristics, health status, disability, marital or family status, conviction, age, appearance, membership in political, trade union, and other organizations, and other real or assumed personal characteristics.[14]

As could be concluded from the definition provided in the mentioned provisions, discrimination can be conducted in any way and can be based on any personal characteristic. The condition that must be met in order to consider an active or passive behavior inequitable is unjustified differentiation among citizens. In the context of this work, the terms 'justified' and 'unjustified' will not be explained. Attention will be paid to the specific manifestation of discrimination through the exercise of freedom of expression. It is related to hate speech.

With the development of the Internet and social networks, opportunities for spreading intolerant, discriminatory, and hateful content have become greater. Bearing this in mind, legal and institutional mechanisms for grappling with communicative aggression on the Internet have been developed during the last decades and within the current century in general.[15] The Law on Public Information and Media from 2014 and the Code of Journalists of Serbia define the media's obligation to edit all content contained in their publications, including readers' comments below the texts on online portals. The media are obliged to beware of and prevent the spread of hate speech and aggressive communication through its publications. Social networks, forums, and other online platforms, however, are not subject to these rules, and it is more difficult to edit discriminatory content and take measures against creators of hate speech.

The fact that hate speech is regulated by several laws illustrates Serbian society's sensibility with regard to the consequences of destructive communication. In that sense, breach of the prohibition of hate speech, as a specific kind of discrimination, is treated differently in comparison to other human rights breaches.

12 Etinski, 2013, pp. 51–67; Etinski, 2005, pp. 325–352.

13 Official Gazette of the Republic of Serbia, No. 22/2009.

14 Mršević, 2014, pp. 13–16.

15 Intensive anti-discriminatory action has been expected after conflicts in the Balkans region. For more on the development of the anti-discriminatory approach: Saša Gajin (ed.), 2010; Zubčević et al., 2017; Reljanović, Matić, and Ilić, 2010; Saša Gajin (ed.), 2015; Rašević, 2014.

Specifically, apart from the criminal offense that includes hate speech as a substantial element of the crime, the protection the general anti-discriminatory law provides should be implemented in the dispute via the court decision.

Specific rules are applied in discrimination disputes. Fault for breach (intention or negligence) is not required for the determination of hate speech. The only claim in which fault is relevant is damage compensation. In that case, the absence of the intention to spread or encourage hate, aggression, or violence can release someone from the liability to compensate for damages. This is not the case in other claims.[16]

Further, regarding civil protection, the essential difference between the procedure for protection against discrimination and hate speech according to the Law on Prohibition of Discrimination and the procedure according to the Law on Public Information and Media[17] is reflected in passive legitimacy, i.e., the person against whom the request is directed. Whereas according to the Law on Prohibition of Discrimination, the protection of the endangered or violated right to equality is implemented via a demand directed against the discriminator, in lawsuits under the Law on Public Information and Media, the demand is directed against the editor-in-chief of the media publication.[18] The editor-in-chief's liability arises from their responsibility for the content they edit. This is regardless of who authored the information, whether they are a journalist or someone else. In other words, the editor-in-chief's responsibility exists not only when the author of the information is a journalist, but also when the hate speech published in the media comes from others. In online media publications, indiscriminate publication of readers' comments on a certain topic in which the ban on hate speech is violated is not rare.[19]

Protection under criminal law is narrower because the criminal offense exists in cases of inciting national, racial, and religious hatred and intolerance.[20] The offender's guilt depends on proving that their intent in this case was aimed at inciting religious hatred and intolerance among nations. To be precise, the fact that there was a verbal or physical attack on another people, nationality, ethnic community, or religion does not mean that the subjective (and substantial) element of this crime has been demonstrated.

16 According to Art. 43 of the Law on Prohibition of Discrimination, through a lawsuit, the plaintiff may demand: 1) imposing a ban on an activity that poses the threat of discrimination, a ban on proceeding with a discriminatory activity, or a ban on repeating a discriminatory activity; 2) that the court should establish that the defendant has treated the plaintiff or another party in a discriminatory manner; 3) taking steps to redress the consequences of discriminatory treatment; 4) compensation for material and non-material damage; 5) that the decision passed on any of the lawsuits referred to in previous items be published.

17 Law on Public Information and Media, Official Gazette of Republic of Serbia, No. 83/2014, 58/2015, 12/2016.

18 Art. 103 of the Law of Public Information and Media.

19 Rašević, 2018, pp. 1309, 1310.

20 Art. 128 of the Criminal Code.

2.3.4. Breach of the right to privacy

Disclosure of personal data, details about private life, photos, and other information without the concerned person's consent violates the right of privacy.[21] This right is a focus of modern legislature, precisely because of the increased incidence of uncontrolled use of gathered personal information. In most cases, it is about abuse.

The increasing popularity of the Internet, especially social networks, has led to a more intensive consideration of privacy protection. Specifically, social network websites contain user information such as age, relationship status, income, and information about close family members, as well as registered users' addresses. The most common abuses and violations of privacy using data available on social networks are identity theft and the manipulation of personal data, which are related to employment and the misuse of photos on the Internet.[22] A large number of social network services store personal data about users, so that users do not have to re-enter them when they want to use them later, e.g., for online shopping, booking over Internet sites, or simply revisiting the same website.[23]

Simultaneously with the development of artificial intelligence and machine learning, expectations regarding the appropriate improvement of the legal instruments pertaining to personal data protection have arisen. Ethical issues arising from the application of artificial intelligence are a special aspect of the human rights debate amidst the so-called Fourth Industrial Revolution. This seems to be even more dominant if the fact that humans spend the majority of their daily lives under a certain type of (in)direct monitoring, intentionally or spontaneously, is considered. Smart devices connected to the Internet, the network of public and private surveillance cameras, and automated decision-making based on online behavior history are just some of the currently recognized sources of dispute between citizens, public policy, and industry.

The misuse of photos on the Internet is a form of privacy violation that occurs when photos from social network users' accounts are used and displayed without their consent. A photo of any person can be displayed in a way that may harm them on a personal level. Over time, social network users could download the photo, resulting in a large number of users having an opportunity to see the photo, share it, or even worse, make a photomontage that could be further misused. In addition, some social networks (e.g., Facebook) note in their terms of use that they reserve the right to publish user information or share it with other companies, lawyers, courts, government agencies, etc., if deemed necessary.[24]

21 Judgement of the Court of Appeal in Belgrade, Gž3 77/18, as of 29 March 2018.

22 Identity theft consists of the unauthorized use of personal data (date of birth, current residence, telephone number, occupation, friends, personal pictures), which have become publicly available. Identity theft on the Internet is a form of fraud committed by computer users involving gathering personal and financial information through a fake e-mail or website. Đukić, 2017, pp. 99–116.

23 Midorović and Sekulić, 2019, pp. 1158–1159.

24 Diligenski and Prlja, 2018, pp. 27–31.

In this context, children are a particularly vulnerable category, and for this reason, wide, decisive action on the part of competent bodies is more than welcome.

3. Controlling Internet content

In order to protect individuals' human rights and the core issues of state and public interest, a degree of control over expression is necessary, in line with the constitutional purpose. However, since the Constitution explicitly prohibits censorship, a further explanation will be given as to what kind of control Serbian law provides.

Censorship, as an administrative procedure that precedes the announcement or publication of information, is uncommon in the modern world, and it is against the fundamental principles of modern human rights in democratic societies. However, abridgement does not always depend on an official government decree, injunction, or licensing decision.[25] Many governmental measures may have the effect of abridging freedom of thought and expression. In this subsection, those regulated by relevant laws will be examined.

3.1. Registration of media

As previously mentioned, freedom of the media allows everyone the option to establish newspapers and other forms of public information, without prior permission. Serbian law mainly regulates public information services and public media. With respect to that, this paper will focus on those that are operable for freedom of thought and expression and freedom of the media.

First, there is the Law on Public Information and Media. According to this act, public information is provided by the media.[26] The aim of the legal provisions is to obtain and protect the presentation, receipt, and exchange of information, ideas, and opinions through the media in order to promote the values of a democratic society, prevent conflicts and preserve peace, disseminate truthful, timely, credible, and complete information, and enable the free development of the individual.[27]

This law regulates the manner of exercising the freedom of public information, which includes, in particular, the freedom to collect, publish, and receive information, the freedom to form and express ideas and opinions, the freedom to print and distribute newspapers, and the freedom to produce and publish audio and audio-visual media services, information, and ideas through the Internet and other

25 Zuckman et al., 1999, p. 72.
26 Art. 1 of the Law on Public Information and Media.
27 Art. 2 of the Law on Public Information and Media.

platforms, as well as the freedom to publish media and conduct public information activities.[28]

The term 'media' is defined as a means of public information that conveys editorially formed information, ideas and opinions, as well as other content intended for public distribution and an indefinite number of users in words, images, or sound.[29] The term 'media' under this law excludes platforms, such as Internet forums, social networks, and other platforms that facilitate the free exchange of information, ideas, and opinions among its members, or any other independent electronic publication, such as blogs, web presentations, and similar electronic presentations; unless they are registered in the Media Register, they are not considered to be media under this law.[30] Consequently, the provisions of the Law of Public Information and Media are not applicable to unregistered entities (i.e., inapplicable to social networks).

From the perspective of the freedom of expression, several advantages could be recognized through registration.[31]

Journalists are not obligated to reveal their information sources, except where the information refers to a criminal act or the perpetrator of a criminal act for which a sentence of imprisonment of at least five years is prescribed by law and if the information cannot be obtained in any other way.[32] Otherwise, such an obligation exists. Furthermore, for some criminal acts, stricter punishment is stipulated if the victim is a journalist associated with registered media.[33] This should

28 This law also regulates the principles of public information, public interest in public information, the provision and distribution of funds for public interest, imprint, abbreviated imprint and identification, the publicity of media data and the Register, the protection of media pluralism, the position of editors, journalists, and representatives of foreign media, media distribution, temporary storage and insight into the media record, special rights and obligations in public information, personal information, the means and procedures of legal protection, supervision over the application of the provisions of the law, and penal provisions. Art. 3 of the Law on Public Information and Media.

29 Art. 29 of the Law on Public Information and Media.

30 Art. 30 of the Law on Public Information and Media. The Media Register is maintained by the Business Registers Agency of the Republic of Serbia (hereinafter, 'the Agency') in accordance with the law governing the legal position of the Agency and the law governing the registration procedure with the Agency and this law (Art. 37 of the Law on Public Information and Media.). Natural and legal persons, domestic and foreign, have equal rights to publish and other rights pertaining to publishing, in accordance with law and signed international agreement (Art. 11 of the Law on Public Information and Media.). In case law: Decision of the Court of Appeal in Belgrade, Gž. 216/18, as of 21 December 2018.

31 Naturally, there are consequences. According to the provision of Art. 44, the Republic of Serbia, Autonomous Province and a local self-government unit, as well as an institution, a company or another legal person whose majority shareholder is the state, or which is entirely or predominantly funded from public revenue may not co-finance projects of or in any other way allocate state aid to a medium or a publisher not entered in the Register. The Republic of Serbia, Autonomous Province and a local self-government unit, or an institution, company or another legal person whose majority shareholder is the state, or which is entirely or predominantly funded from public revenue may not advertise in or use other services of unregistered media.

32 Art. 52 of the Law on Public Information and Media.

33 Art 138 of the Criminal Code.

act preventively, since strong punishment could deter potential perpetrators. On the other hand, if journalists associated with registered media have committed such a crime, they could be released of punishment if they believed in the truthfulness of the distributed information.[34] Other perpetrators would be punished even if they prove that the information is true.[35]

Within the area of civil law, editors and journalists have an obligation to check the information's origin, truthfulness, and completeness before releasing a publication that contains information about a certain phenomenon, event, or person. Apart from the fact that the journalist must be proven to be at fault, unlike the presumption under the general rules of tort law, all other aspects can be considered stricter than the general rules of damage compensation. In that sense, and because of this, an editor-in-chief is responsible for that medium. An editor responsible for a specific issue, section, or program unit shall be responsible for the content they edit.[36] This liability is not just for content that the editor chooses and directly controls, but also for readers' comments as well.[37]

Apart from the special treatment of registered media, the Law of Public Information and Media includes provisions that specifically protect media pluralism. One of the relevant issues is the Media Register, since its purpose is to provide the public with information about the media.[38] It seems justified to enable citizens to form their own opinions about the authenticity and reliability of information, ideas, and opinions published in the media to facilitate the identification of the media's possible influence on public opinion and protect media pluralism.[39] However, the provisions pertaining to the protection of media pluralism are more important.[40]

In that sense, a threat to media pluralism in the case of printed media shall be identified by the ministry responsible for information, and if there is the merging or cross-acquisition of shares, where at least one electronic medium is involved, the

34 Art. 173-175 of the Criminal Code.

35 Art. 172 of the Criminal Code.

36 Art. 48 of the Law on Public Information and Media.

37 Judgement of the Court of Appeal in Belgrade, Gž3 137/19, as of 21 August 2019.

38 Art. 38 of the Law on Public Information and Media.

39 Art. 7 of the Law on Public Information and Media. Therefore, media registration does not depend on the decision process in terms of the authorities' jurisdiction, but rather in terms of public information. Hence, registration can only be rejected if the application does not fulfil the formal conditions (Art. 17 Law on the Procedure of Registration in Business Register Agency, Official Gazette of Republic of Serbia, No. 99/2011, 83/2014 и 31/2019.). Otherwise, the registrar will register the media. According to available data, there are 2,642 active media in Serbia. Available at: https://bit.ly/3kp5oxc. Registration, however, is not necessarily permanent. Media should be deleted following a publisher's notice, or the registrar, acting in an official capacity, shall delete the medium from the Register based on the decision of the responsible authority referred to the protection of media pluralism, following the deletion of the publisher from the Register where it was entered into or for any other reason stipulated under a special law (Art. 41 Law on Public Information and Media). See more on this topic: Rakib, 2014, pp. 337–349.

40 Art. 6, and 45-47 of the Law on Public Information and Media.

threat shall be identified by an independent regulatory body, in accordance with the law regulating electronic media.[41] The ministry responsible for information shall initiate the procedure, and whenever it is established that media pluralism has been threatened, the strictest measure that can be levied on an entity is the medium in question's deletion from the Register. In that case, the entity loses its media status and consequently is not in the capacity of the Law on Public Information and Media.[42]

When it comes to electronic media, the situation regarding controlling media is different. Specifically, the Law on Electronic Media[43] stipulates, in accordance with international conventions and standards, the organization and operation of a regulatory body for the electronic media, conditions and the manner of providing audio and audiovisual media services, conditions and procedures for issuing licenses for the provision of audio and audiovisual media services, and other issues relevant to the field of electronic media.[44] Social networks are not covered by its provisions.

The regulatory body for electronic media is an independent regulatory organization, that is, a legal entity that exercises public authority for the purpose of, among others, contributing to the preservation, protection, and development of freedom of opinion and expression.[45] Within the scope of its work, the regulator controls media service providers'[46] operation and ensures the consistent application of the provisions of the relevant law, imposes measures upon media service providers, and decides on complaints in connection to media service providers' programming activities.[47]

Examining controls on media service providers' operation, the regulator observes whether such operation reflects the consistent implementation and improvement of the principles underlying the regulation of relations in the field of electronic media and immediately applies the required measures. Such control pays special attention to media service providers' compliance as it regards the program content they are licensed to broadcast. The regulator, before a competent court or other public authority, shall initiate proceedings against the media service provider

41 Art. 47(1) of the Law on Public Information and Media.

42 Moreover, a fine of between RSD 100,000 and 1,000,000 for a commercial offense shall be imposed on a legal person—a publisher who does not act as cautioned by the competent body in proceedings establishing that media pluralism has been threatened. A fine of between RSD 10,000 and 200,000 for a commercial offense referred to as previously mentioned shall be imposed on a responsible person in the publisher's company (Art. 133 of the Law on Public Information and Media).

43 Official Gazette No. 83/2014 and 6/2016.

44 Art. 1 of the Law on Electronic Media.

45 Art. 5(1)(2) of the Law on Electronic Media.

46 Media service provider means a natural or legal person who has editorial responsibility for the choice of audiovisual content of an audiovisual media service, (i.e. audio content of a radio media service) and determines the manner in which it is organized. Art. 4(1)(6) of the Law on Electronic Media.

47 Art. 22 (4-11) of the Law on Electronic Media.

or the person responsible if their act or omission has the character of an offense punishable by law.[48]

Natural and legal persons, including media service providers, are eligible to submit applications to the regulator in relation to program content if they believe that the content is violating or jeopardizing their personal interests or the public interest.

The procedure for applications against media service providers is regulated by the Law on Electronic Media and detailed bylaws. What is important in the context of freedom of expression are the measures the regulator can take if it is determined that the application is reasonable. Specifically, the regulator can impose on the media service provider a remonstrance, warning, temporary ban on the publication of the program content, or may revoke their license due to a violation of obligations related to the program content,[49] as well as due to a violation of the conditions set forth in the license or approval for providing media services. Those measures should be imposed independently of the use of other means of legal protection available to the injured or another party in accordance with the provisions of special laws. Since the regulator exercises public authority, it is of great importance to preserve objectivity, impartiality, and proportionality. Hence, during the process of imposing measures, the registrar is obliged to allow the media service provider to comment on the facts pertaining to the reason for the procedure. Finally, the registrar can submit a request for misdemeanor and/or criminal proceedings or initiate other proceedings before the competent state body and refer the applicant to how they can achieve and protect their rights.[50]

As can be concluded from the aforementioned, the registrar has the potential to intervene in the content that can be distributed through the electronic media. However, that intervention is not censorship because it comes after the law has been breached. Moreover, if the application of the relevant provisions in practice is taken into account, it can be said that the registrar has not used their potential to the extent to which one would expect. In other words, this can provide electronic media content that is much more acceptable to the public interest of freedom of expression and freedom of information.[51] Moreover, it can be said that the main problem in the registrar's power is room for bias, since the registrar has failed to react adequately to each application.[52] If the fact that the registrar's decisions and measures are under the review of the public and the court is considered,[53] this problem seems to be smaller in practice.

48 Art. 24 of the Law on Electronic Media.

49 Defined in Articles 47-71 of the Law on Electronic Media.

50 Art. 26 of the Law on Electronic Media.

51 Conclusion based on decisions of the Regulator, available at: http://www.rem.rs/sr/odluke/.

52 Similar observation has been stressed in the Strategy for the Development of Public Information in the Republic of Serbia for the period 2020–2025, prepared by the Serbian Government, Official Gazette of the Republic of Serbia, No. 11/2020.

53 Art. 38-42 of the Law on Electronic Media.

3.2. Providing the infrastructure

The provisions analyzed in the previous section are in relation to the media under Serbian jurisdiction.[54] As regards media service providers that do not belong under the jurisdiction of Serbian law and authorities, especially social networks that are not regulated by state law, it seems impossible to control the content that is distributed among users and the public in a wider sense.

However, it is not possible to upload and distribute content on the Internet, regardless of the platforms where the information is accessible, without the physical layer of the Internet, that is, its infrastructure.[55] Given those parts of the Internet infrastructure that are located within a state's territorial borders, there is only a segment of the global network over which full control can be established. By virtue of the state's sovereignty over the territory, it has certain power to regulate Internet infrastructure. This mainly derives from the regulation of electronic communications,[56] but also from other areas.[57] In this way, the state is also competent to indirectly regulate the Internet and the flow of data within its territory. Via regulations in the field of electronic commerce, the state regulates information society services, which are

54 A media service provider is under the jurisdiction of the Republic of Serbia if: 1) it is established in the territory of the Republic of Serbia; 2) it is not established in the territory of the Republic of Serbia, but: a) it uses a terrestrial satellite transmitting station that is located in the Republic of Serbia, and/or b) it uses satellite capacity appertaining to the Republic of Serbia. A media service provider shall be deemed to have been established in the Republic of Serbia if: 1) its head office is located in the Republic of Serbia, and its editorial decisions about media services are made in the Republic of Serbia; 2) its head office is located in the Republic of Serbia, and its editorial decisions about media services are made in another member state of the European Union, provided that a significant number of persons are employed in the Republic of Serbia (under contract of employment or otherwise) and are involved in carrying out activities related to media services; 3) its head office is located in the Republic of Serbia, and a significant number of persons employed under contracts of employment or otherwise involved in carrying out activities related to media services work in the Republic of Serbia and another member state of the European Union; 4) it initially commenced its activity—in accordance with the law—in the Republic of Serbia, under the condition that it maintains a stable and effective relationship with the Serbian economy and that a significant number of persons—employed under contracts of employment or otherwise involved in carrying out activities related to media services—do not work in one of the member states of the European Union; 5) its head office is located in the Republic of Serbia, and its decisions about media services are made in a country that is not a member state of the European Union, or vice versa, under the condition that a significant number of persons—employed under contracts of employment or otherwise involved in carrying out activities related to media services—work in the Republic of Serbia. If it cannot be determined whether a media service provider is under the jurisdiction of the Republic of Serbia or any other member state of the European Union, the media service provider shall be under the jurisdiction of the member state in which it was established within the meaning of Articles 56 through 58 of the Stabilization and Association Agreement concluded between the Republic of Serbia and the European Communities and their member states (Art. 45 of the Law on Electronic Media).

55 It includes, basically, cables, routers, servers, and other telecommunications devices.

56 The Law on Electronic Communication, Official Gazette of the Republic of Serbia, No. 44/2010, 60/2013, 62/2014 and 95/2018.

57 For example, laws on national security, defense, etc.

usually provided for a fee, remotely and by electronic means, upon the request of the recipient of the service. As hosting services represent information society services, regulations in the field of e-commerce apply to the hosting company.[58]

3.2.1. Jurisdiction over electronic communications

According to the Law on Electronic Communications:

> The activity of electronic communications is a regulated activity that includes the construction or installation, maintenance, use and provision of public communications networks and the provision of publicly available electronic communications services. Electronic communications network is transmission systems and switching and routing devices that enable the transmission of signals by wired, radio, optical or other electromagnetic means, including satellite networks, fixed and mobile networks. Electronic communications service is a service that is generally provided for a fee and consists of the transmission of signals in electronic communications networks, including telecommunications services and services for the distribution and broadcasting of media content. Outside the field of electronic communications, in the field of construction, the state also regulates the conditions for setting up the physical infrastructure necessary for performing the activities of electronic communications.[59]

Without physical infrastructure in a certain state's territory, there is neither a network nor the possibility of access. Since physical infrastructure cannot, at least legally, exist without approval, permits, and consent from state bodies, the state is able to indirectly influence other layers of the Internet through the physical one.

States most often apply their general regulations to the Internet, to the extent that their control mechanisms allow. The state has *de facto* control over servers, which are like computers located in its territory, where data available to the global network are stored.

Public authorities in certain situations have the authority to access and take possession of servers. As the servers are usually not owned by the persons who store data on them, but rather by private hosting companies that rent these servers to interested parties, the system of state control is reflected through regulatory competence over these companies.

Electronic surveillance is traditionally linked to state bodies such as security, internal affairs, and defense services. Those bodies, to which the highest state interests are entrusted, use operators' telecommunication means and information systems to intercept communication and access communication data. Bearing in mind that this

58 The Law on Electronic Commerce, Official Gazette of the Republic of Serbia, No. 41/2009, 95/2013 and 52/2019.

59 Art. 4 of the Law of Electronic Communications.

is a serious encroachment on citizens' right to privacy, the Constitution and laws provide for procedural guarantees and provisions that protect against abuse. The regulation that most thoroughly deals with electronic surveillance in Serbia is the Law on Electronic Communications.

Legal interception of electronic communications is the secret surveillance of telecommunication operators' electronic communications services, activities, and traffic; it is related to the content of communications and performed by authorized state bodies or organizations.[60] On the other hand, the Law on Electronic Communications introduces the obligation to retain data, as a result of which each operator is obliged to keep data on communication for a period of one year. These data do not refer to content, but rather to the type of communication, its source, destination, beginning, duration, and end, as well as data on the device through which the communication has been performed and the location of that device, so that state authorities can access them in cases provided by law. When it comes to the Internet, this practically means that all Internet communication operators are obliged to keep a whole range of data that they can collect when viewing each individual packet moving through their network, without intruding on the content of the communication.

Interception of communication and access to communication data are allowed only for a certain period of time and on the basis of a court decision, if necessary for the purpose of conducting criminal proceedings or protecting the Republic of Serbia's security, in the manner stipulated by law.[61]

In the described manner, Serbian authorities can achieve indirect control over the global network. Communication content is accessible to state bodies, at least in a limited number of cases. Moreover, access to Internet use data is unlimited (e.g., who, when, and at what location they accessed the network).

3.2.2. Jurisdiction over Internet platforms

The creator of an online platform or website's content, as well as the platform itself, can be kept on servers located in a territory different from the state to which they belong or the state from which the citizens who are the object of the content originated. Apart from that, the non-national domain is also available. This means that creators can maintain their anonymity through a registered domain system. For that reason, states may face significant obstacles in the identification process. Hence, control over content can be achieved only at the infrastructure and logical levels.[62]

In cases where the platform is hosted on a server outside state territory, the questions of the hosting provider's cooperation with foreign state bodies and the

60 Art. 127 of the Law of Electronic Communications.

61 Art. 130 of the Law on Electronic Communication. See also: Reljanović, 2015, pp. 113–124.

62 For example, by ordering the hosting provider to remove the content, which is not a particularly effective measure due to the fact that a skilled Internet user can re-install the removed content on a large number of servers, as well as on servers in a different jurisdictions.

fulfillment of their requirements are also raised. Since they do not have any mechanisms to directly remove the disputed site or content or undertake some other measure in accordance with their own rules, state actions rely on international legal assistance instruments.

In cases of limited international cooperation, this can create insurmountable problems for states seeking to remove Internet content that is illegal within their jurisdiction. If the state intends to punish the platform's owner/author for violating domestic regulations, such a possibility exists if the latter has a registered or representative office in its territory.[63]

3.2.3. The Internet service provider's liability in Serbian law

An Internet service provider or intermediary provides a service connecting entities that provide information with those requesting that information. Services can have different content and are most often differentiated into three groups: mere conduit, caching, and hosting.

The most common problem emerges on a daily basis when the Internet provider enables the posting of illegal content and its sharing among users.[64] In the context of this paper, this service is of great importance.

The provider does not directly commit the injury, but by providing the service, it enables the injury. The service provider's responsibility is, therefore, indirect (shared or liability of another). According to the general rules on liability, the determination of liability is grounded in the issue of conscientiousness, i.e., the question of whether the Internet provider knows or may know that copyright infringement is committed while providing the service. In other words, the provider's responsibility depends on the question of whether it applies fair trade principles while providing the service.

The technological environment in which data are exchanged, the amount of information transmitted via the Internet, and the speed of their flow are only some of the circumstances that make the assessment of conscientiousness difficult. Consequently, it is rather difficult to distinguish situations in which it has been assumed that the provider knows or could have known about the rights violation. This is especially important if the general absence of the Internet service provider's rights and obligations to supervise, i.e., control communication among users, is taken into account.

The liability of Internet intermediaries, i.e., information society service providers under the law of the Republic of Serbia, is normally regulated by the Law on Electronic Commerce.

63 The list of representatives of the biggest foreign companies that collect Serbian citizens' personal data in the course of their business activity is available at: https://predstavnici.mojipodaci.rs/. The most popular digital platforms, e.g., social networks including Facebook, Instagram, WhatsApp, and Twitter, have not identified a representative in Serbia to do that, even after several years of initiative.

64 Radanovich, 2016, pp. 157–160.

According to this law, the intermediary is liable for the violation when it knows or could have known about service users' unauthorized actions or the content of the data and does or has not removed or disabled access to the given data immediately upon acknowledgement of an unauthorized action or data.

The obligation to monitor the content stored and exchanged via the Internet is not prescribed. However, if there is reasonable doubt that illegal actions and the exchange of illegal content are being performed by using the service, the provider is obliged to inform the competent state authority. Disclosure of user data, content removal, and disabling access to suspicious content are possible only based on a court or administrative decision. The latter directly complements the issue of liability for damage compensation. If it is assumed that guilt, as a condition of liability for damages, can be excluded when the intermediary proves that they did not act with intent or negligence, the crucial fact depends on what is considered to be negligence in providing services.

It can be imagined that the provider would prevent further infringement by disabling access to or removing unauthorized content. However, such actions can violate other entities' rights. Since the provider, in principle, has no obligation to supervise the exchanged content, the very notification of the existence of a breach instigates the provider's obligation to alert the competent authorities and act in accordance with their decisions.

In the context of liability for damages, the provider is considered liable when it has received notification of a possible infringement and fails to alert the competent authorities. Additionally, the provider can be considered liable even if, on the basis of an appropriate judicial or administrative act, it fails to provide information that is relevant for the detection of the person whose action directly caused the damage. In both cases, it is about the provider's guilt arising from its failure to act with the expected care.[65]

3.2.4. Self-regulation of social networks

In order to achieve the standards of 'good practice' and 'fair trade' when providing a networking service and, in particular, to secure the service within the safe harbor principles, social network platforms include measures and procedure in their terms of use that are applicable on the occasion of rights violations. Bearing in mind that this chapter focuses on Serbian law and given that there are no such platforms under Serbian jurisdiction (or, at least, no popular and widespread ones), the major obstacles to protecting freedom of expression, as well as rights violated by the exercise of the freedom of expression, must be mentioned.

Specifically, social network user accounts do not always identify the user who posted the content on the network. When this content is harmful, the person whose right has been violated cannot get any information about the wrongdoer without

65 Radovanović, 2015, pp. 85–87.

a decision from state authorities. In Serbia, this would mean that the Ministry of Internal Affairs should apply for international legal help in order to get the necessary data. According to the relevant United States law, such data can be obtained only when the security of a state or its citizens is endangered. In addition, when a company blocks a suspect account, although that can be helpful in some cases, the circumstances under which the decision has been made are outside of the frame of legal principles; that is, the procedure is self-organized and not transparent. Furthermore, the issues upon which the company has decided are, at their core, under the court's jurisdiction. All these circumstances leave room for bias, particularly in the sense of the unfounded rejection of the right to expression and information, i.e., information exchange.

4. Concluding remarks

Having observed the laws regarding freedom of expression, its restrictions, and the exchange of information, the main legal difference between social networks and media can be noted. The media are responsible for all content published in their publications, even readers' comments that are published below the texts on their online portals. The media are considered to have enough capacity to review and approve all such content. Social networks, *de facto*, do not and cannot have the capacity (mainly for legal reasons) to view every user post. This seems more than acceptable when it comes to freedom of expression, but it is problematic when it comes to its restriction.

Although according to the valid law of the Republic of Serbia, social networks are not media, their importance in Serbia's information sphere and on the media scene in general is becoming more pronounced. Traditional media often use influential tweeters, well-known bloggers, and Facebook statuses that have achieved high readership and been widely shared as relevant information sources, and posts on social networks are transferred to traditional media and commented on. Moreover, when celebrities comment on current socio-political circumstances, their social media publications become regular content in traditional media.

Social networks enable fast communication and information exchange. Social networks are premised upon interaction with and the expression of one's own views, comments, and contributions to discussions to other social media participants. Indeed, the term 'social networking' implies user interaction. Electronic propaganda flows uninterruptedly, since it is easier to reach the target group of users interested in certain information on social networks. Many personal profiles on social networks that are filled with certain content or information, as well as reactions to current social events, no matter how subjective, play a significant role in creating social events.

Traditional media in Serbia still take the lead in terms of informing and creating public opinion and social trends. However, the increasingly frequent reactions of traditional media to events on social networks indicate the latter's gradual influence and their inevitable inclusion in media flows. However, social networks are not (at least, directly) encompassed under Serbian law and jurisdiction.

The direct enforcement of domestic law on the Internet is possible only where the state has sovereignty, which is expressed through territorial and personal authority over certain segments of architecture and content.[66] The only way for the state to fully implement its legal system on the Internet is to take full control of the physical and logical layers. This would be a more than ambitious attempt since the Internet is a very complex environment that is further complicated by the fact that many companies providing various services within the Internet's physical and logical infrastructure are based in other countries.

On the other hand, the law leaves the possibility for the platforms themselves to remove the content that they assess as inappropriate or offensive. This means that companies cannot be accused of restricting the constitutionally guaranteed rights to freedom of speech as long as their regulation has been adequately targeted. Companies have taken advantage of this opportunity to develop algorithms for recognizing offensive or inappropriate content and eliminating it quickly. Algorithms have defects because machines still cannot replace a human's approach to editing content. This provides an additional argument that should prompt the authorities to ask for stricter rules for the regulation of social networks.

5. Impact of 'fake news' on freedom of expression

Misinformation concerning the politics, economics, health, and other societal spheres is probably as old as society itself. In the era preceding the media, particularly the Internet, this problem was in focus within small groups. At present, given that global network communication has intensified information exchange, making it easier and faster, the exponential increase of fake news demonstrates its potential to harm or at least endanger fundamental human rights. The fake news phenomenon has reached a new level worldwide. As such, it has been the subject of research in various areas, including law. The Serbian legal system's approach to the fake news phenomenon will be analyzed in this section. Satirical content, parody, and similar legally protected 'false' speech will be put aside in this analysis.

66 Though it is very rare in practice, the application of the Serbian law can be through the provisions of the competent law in the collision of laws. Popović and Jovanović, 2017, pp. 35–63.

5.1. Meaning of 'fake news'

'Fake news' in its pure meaning warns about deception. It is the opposite of truth, reality, and fact-grounded statement. However, to simplify the global challenge of fighting fake news, the distinction between this phenomenon and disinformation as such is eliminated. The dangerous element of fake news arises from the fact that it is 'intentionally and verifiably'[67] false and could impact readers, sometimes with serious consequences. For the purpose of this work, the term 'fake news' will be used to refer to the online publication of intentionally or knowingly false statements of fact.[68]

The major problem with fake news lies with its qualification as news. The attribute 'fake' is only subsequently ascribed, often when readers have already built their opinion or taken unlawful action (e.g., left a comment that discriminates against someone mentioned in the false story).[69] There is no technical or legal means to prevent the distribution of such information. The focus is therefore on the reaction.

5.2. Regulation of fake news in Serbia

The social harmfulness of spreading false news in Serbia was recognized decades ago. Accordingly, the Criminal Code once contained a crime called *spreading false news*. This crime is committed by anyone who spreads fake news or rumors that they know to be false, with the aim of causing a serious violation of public order and peace. A person would be fined or imprisoned for up to one year for this crime, if the crime was committed through the press, radio, television, other similar media, or at a public gathering. Although the crime in this form has ceased to exist, that does not mean that spreading false news is now permissible.

5.2.1. Criminal law concerns related to fake news

The criminal offense has changed and has existed in the changed form since 2006 in Article 343 of the Criminal Code. The crime is entitled *causing panic and disorder*. This criminal offense is committed by a person who, by presenting or spreading false news or allegations, causes panic or serious disturbance of public order or peace, or thwarts or significantly hinders the implementation of decisions and measures of state bodies or organizations exercising public authority.

67 Allcott and Gentzkow, 2017, p. 213. According to this author, fake news rules out several causes: 1) unintentional reporting mistakes, 2) rumors that do not originate from a particular news article, 3) conspiracy theories (these are, by definition, difficult to verify as true or false, and they are typically originated by people who believe them to be true), 4) satire that is unlikely to be misconstrued as factual, 5) false statements by politicians, and 6) reports that are slanted or misleading but not outright false (p. 214).

68 Klein and Wueller, 2017, p. 6.

69 Colliander, 2019, p. 204.

This criminal offense is punishable by imprisonment from three months to three years and a fine, and if the criminal offense is committed through the media or similar media or at a public gathering, a prison sentence of six months to five years is threatened. As the consequence of presenting or spreading false news, the legislator, in addition to disturbing public order and peace, also foresaw the frustration or significant obstruction of the decisions and measures of state bodies or organizations exercising public authority, and the penalties have been increased.

This does not mean that a journalist, when publishing news, must establish the absolute truth. If there was an obligation to establish the absolute truth before news could be published, it would deter journalists from publishing texts for fear of possible criminal prosecution if it happens that the news they published is not true.

Although it is not necessary to establish the absolute truth, it certainly does not release journalists from the duty to check the truthfulness of the information as required by the rules of the journalistic profession and the Code of Journalists of Serbia. This crime can be committed only with intent, which would mean that the journalist, at the time of publishing the news, has to have been aware that the information was false. If the journalist, before publishing, checks the truthfulness of the information in the manner prescribed by the Law on Public Information and Media and the Code of Journalists of Serbia, there is no awareness that the information being published is false, and therefore, this crime does not exist.

In order for this crime to exist, it is not enough that the information published is false; its publication must cause the consequences provided by the Criminal Code: panic or serious disturbance of public order or peace, or significant hindrance of the implementation of state organs' decisions and measures.

The Criminal Code stipulates that a person who transmits false news through the media or similar media or at a public gathering commits a more serious form of this crime because the false news will then be available to a larger number of people. Therefore, a more severe punishment is envisaged for the more serious form of this criminal offense. If a fake news item is published through a newspaper, radio, television, or other public medium, and the author of the fake news item is unknown or the fake news item is published without the author's consent or there are legal obstacles to prosecuting the author, as a rule, the responsible person is the editor-in-chief of that medium. If these circumstances exist on the author's part, and the false news is published in an occasional printed publication, its publisher will be responsible. Under the same conditions, the manufacturer will be responsible.

5.2.2. Civil law concerns related to fake news

In order to protect public interest through the general and special prevention of future violations, criminal law threatens severe punishment. However, absence of the crime does not mean that human rights are not being violated. Apart from the fact that crime protection does not exclude the measures other legal areas implement, civil law provides more claims than is the case with criminal provisions.

Based on Serbian case law, it seems that the most frequently invoked claim against fake news creators is for monetary damages. However, unlike the specific rules on damage compensation when a journalist publishes harmful content,[70] when making the same request against the media due to damages caused by fake news, the general rules of the Law on Contracts and Torts are applied. The reason lies in the provisions of Art. 112 and Art. 113 of the Law of Public Information and Media, stipulating the journalist's and editor-in-chief's liability for damages.

Specifically, a person referred to in information that was prohibited to be published in accordance with this law and who suffers damages because of the publication of the information is entitled to damages to cover material and nonmaterial damages pursuant to the general regulations and provisions of this law, disregarding other means of legal protection available to said person in accordance with the provisions hereof. The right to compensatory damages also pertains to a person whose replay, correction, or other information was not published, although its publication was ordered by the competent court when that person suffered damages. Fake news is not forbidden. Hence, this specific law's liability rules do not apply.

Therefore, according to the relevant provision of the Law on Contracts and Torts, a person who causes damage to another is obliged to compensate for it, unless they prove that the damage was not caused through their own fault.[71] In this respect, the Law of Public Information and Media provides an obligation of journalistic due diligence. This obligation requires that prior to publishing information about an occurrence, event, or person, both the editor and the journalist shall check its origin, authenticity, and completeness with due diligence appropriate for the circumstances. Furthermore, both the editor and the journalist shall convey the accepted information, ideas, and opinions authentically and fully, and if the information is taken from another medium, they shall credit that medium.[72] In the context of fault for damage, it is hard to imagine circumstances under which the journalist could be excluded from liability.

The basis of liability is therefore the presumed fault of the person whose action caused the damage. Hence, a person who has undertaken an injurious action against someone's concrete human rights (dignity, honor, reputation) is obliged to compensate for the damage caused by their wrongdoing. However, this general rule faces enforcement obstacles with regard to fake news damages.

Primarily, there are difficulties when it comes to spreading fake news via the Internet. From the aspect of general rules, in order to succeed with a request for compensation, the damaged right holder must prove that there is a causal link between

70 Art. 112-118 of the Law on Public Information and Media.

71 Art. 154(1) Law on Contracts and Torts.

72 Art. 9. In case law: Judgement of the Higher Court of Pančevo, P.br. 7/10 as of 3ª February 2010, Judgement of the Higher Court of Belgrade, P3 244/14, as of 14 March 2016, Judgement of the Supreme Court of Serbia, Rev. 1477/05, as of 27 December 2005, Judgement of the Court of Appeal in Belgrade, Gž 308/2016, as of 3 February 2017, Judgement of the Higher Court in Valjevo, Gž 507/19, as of 14 May 2020, Judgement of the Court of Appeal in Novi Sad, Gž. 83/11, as of 18 January 2011.

the false information and the damage itself. It is obvious that their position in the Internet environment where the injury occurred is far from simple. This is particularly the case when it comes to determining the origin of the misinformation. Apart from that, it is not necessary for the original source of information to be correlated with the damage caused. This could be the case when fake news is artistic, satirical, parody, or some other kind of allowed expression that is posted on a social network platform and reposted on media websites. Finally, regarding monetary compensation for non-pecuniary damages, the success of the demand, such as with respect to amount, depends on the intensity and duration of the suffering.

5.2.3. Concluding remarks

The above observation considered Serbian laws that could be applicable in cases involving fake news that causes legally relevant consequences. The complexity of the fake news phenomenon makes it challenging for legislatures to deal with in terms of exercising control.[73] Fake news is the product of human beings, and society should deal with it through various fields of action.

In this respect, civil action can be of as much importance as authorities' actions. A major part of this is undertaken by different communities and associations that are devoted to identifying fake news, tracing its origin, and investigating liability.[74] Fake news could also be suppressed through increased efforts to educate journalists, as well as through the conscience of the society itself.

From the legal point of view, it would be very problematic to define who should be the arbiter of truth.

73 During the novel coronavirus pandemic crisis in 2020, the government adopted the Conclusion on informing the population about the condition and consequences of the infectious disease COVID-19 caused by the SARS-CoV-2 virus (Official Gazette of the Republic of Serbia, No. 48/2020). The Conclusion was considered an attempt to restrict freedom of expression to the extent that it is disproportionate to the officially proclaimed goal. Thanks to the reaction of the journalistic and media community in Serbia, the controversial conclusion ceased to be effective within a short period of time.

74 For example, in Serbia: https://fakenews.rs/, where everyone could ask for the verification of the information published by the media.

Bibliography

ALLCOTT, H., GENTZKOW, M., (2017) 'Social Media and Fake News in the 2016 Election', *Journal of Economic Perspectives*, 31(2), 211–236.

COLLIANDER, J. (2019) '"This is fake news": Investigation the role of conformity to other users' views when commenting on and spreading disinformation in social media', *Computers in Human Behavior*, Vol. 97, pp. 202–215.

DILIGENSKI, A., PRLJA, D. (2018) *Fejsbuk, zaštita podataka i sudska praksa*, I izdanje, Beograd, Institut za uporedno parvo.

ĐUKIĆ, A. (2017) 'Krađa identiteta – oblici, karakteristike i rasprostranjenost', *Vojno delo*, 2017/3, pp. 99–116.

ETINSKI R. (2005) 'Doprinos Suda evropskih zajednica razvoju zabrane diskriminacije u oblasti osnovnih sloboda', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 2005/3, pp. 325–352.

ETINSKI, D. (2004) *Međunarodno javno pravo*, II izdanje, Novi Sad: Pravni fakultet Univerziteta u Novom Sadu.

ETINSKI, R. (2013) 'Načelo jednakosti i nediskriminacije u pravu Republike Srbije i pravu Evropske unije: uporedna analiza', in Keča, R. (ed.) *Harmonizacija srpskog i mađarskog prava sa pravom Evropske unije*, pp. 51–67.

GAJIN, S. (ed.) (2010) *Anti-discriminatory Laws – Guidelines*, I izdanje, Beograd, Centar za unapređenje pravnih studija.

GAJIN, S. (ed.) (2015) *Diskriminacija u Srbiji 2014 – izveštaj koalicija protiv diskriminacije*, I izdanje, Beograd, Centar za unapređenje pravnih studija.

KLEIN, D.O, WUELLER, J.R. (2017) 'Fake News: A Legal Perspective', *Journal of Internet Law*, 2017/10, pp. 5–13.

KRŠIKAPA, M. (2008) 'Naknada štete zbog povrede ljudskih prava u praksi evropskog suda za ljudska prava', *Bilten sudske prakse Vrhovnog suda Srbije*, 2008/2.

MIDOROVIĆ, S., SEKULIĆ, M. (2020) 'A New Function of Personal Data in the Light of the Contract for the Supply of Digital Content and Digital Services', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 53(3), pp. 1145–1166.

MIJOVIĆ, L.S. (2020) 'Internet i sloboda izražavanja u praksi Evropskog suda za ljudska prava', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 54(3), pp. 1023–1042.

MRŠEVIĆ, Z. (2014) *Zločin mržnje, govor mržnje i grafiti mržnje*, I izdanje, Beograd, Kancelarija za ljudska i manjinska prava.

ORLOVIC, S. (2019) *Ustavno pravo*, I izdanje, Novi Sad, Pravni fakultet Univerziteta u Novom Sadu.

PETRAŠINOVIĆ, S. (2020) 'Fejsbuk I uvreda', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 54(1), pp. 469–485.

POPOVIĆ, D.V., JOVANOVIĆ, M. (2017) *Pravo Interneta – Odabrane teme*, I izdanje, Beograd, Pravni fakultet Univerziteta u Beogradu.

RADOVANOVIĆ, S. (2015) 'Građanskopravna odgovornost internet posrednika za povredu autorskog prava – uporednopravni aspekt', in: *Intelektualna svojina i Internet* (ed. Popović, D.), Beograd, Pravni fakultet Univerziteta u Beogradu, pp. 113–124.

RADOVANOVIĆ, S. (2016) 'Autorsko pravo i Fejsbuk, s posebnim osvrtom na pravo Srbije', in: *Intelektualna svojina i Internet* (ed. Popović, D.), Beograd, Pravni fakultet Univerziteta u Beogradu, 139–164.

RAKIĆ, I. (2014) 'Zaštita medijskog pluralizma u Srbiji prema novim zakonskim rešenjima', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 48(3), pp. 337–349.

RAŠEVIĆ, Ž. (2014) *Zabrana diskriminacije u privatnom pravu*, I izdanje, Beograd, Zadužbina Andrejević.

RAŠEVIĆ, Ž. 'Nediskriminacija u prednacrtu Građanskog zakonika Srbije', *Zbornik radova Pravnog fakulteta u Novom Sadu*, 52(3), pp. 1297–1315.

RELJANOVIĆ, M. (2015) 'Zaštita autorskog prava i pravo na privatnost elektronskih komunikacija u Republici Srbiji', in: *Intelektualna svojina i Internet* (ed. Popović, D.), Beograd, Pravni fakultet Univerziteta u Beogradu, 113–124.

RELJANOVIĆ, M., MATIĆ, M., ILIĆ, G. (2010) *Priručnik za primenu antidiskriminacionog zakonodavstva*, I izdanje, Beograd, Udruženje javnih tužilaca i zamenika javnih tužilaca Srbije.

ZUBČEVIĆ, A. R., BENDER, S., VOJVODIĆ, J. (2017) *Regulatorna tela i govor mržnje*, I izdanje, Beograd Savet Evrope, Regionalna publikacija.

ZUCKMAN, H.L., CORN-REVERE, R.L, FRIEDEN, R.M., KENNEDY, C.H. (1999) *Modern Communication Law*, 1st edn. Minnesota: West Academic Publishing.

Chapter V

Social Media, Freedom of Expression, and the Legal Regulation of Fake News in Croatia

Davor Derenčinović

1. Introduction

According to some estimates, in 2019, slightly more than 50% of the world's population had access to the Internet, while in 2009, this number was significantly lower (less than 5%).[1] Just over a year later, that number was estimated at over five billion people, representing somewhere around 65% of the world's population.[2] In Croatia in 2019, 79% of citizens used the Internet. For comparison, in 2009, 51% used the Internet, and in 1999, only 4% did. This means that in just twenty years, there has been a 75% increase in the Croatian population with Internet access.[3] Some other sources point to as many as 92% of Croatian citizens who are Internet users.[4] When it comes to social media, according to some estimates, in 2017, there were about 2.86 billion users, 3.6 billion in 2020, and it is estimated that by 2025, about 4.41 billion people will have profiles on social networks.[5] According to a marketing agency survey,

1 See: https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2017&start=1990.
2 Ibid.
3 Ibid.
4 See: https://www.internetworldstats.com/stats.htm.
5 Supra note 1.

Davor Derenčinović (2021) Social Media, Freedom of Expression, and the Legal Regulation of Fake News in Croatia. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 141–174. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_5

at the beginning of 2019, "There were 1,900,000 Croats on Facebook, while there were about 1,100,000 on Instagram."[6] This exponential growth of users has not been coupled with increased media literacy, knowledge about the risks of victimization on global networks, or public awareness about harmful content.[7] Legal regulation of the Internet and social networks is a problem that has just recently come under attention and which raises many questions, including: What are the (potentially) harmful side effects of freedom of expression on the Internet? How can we protect those who are the most vulnerable from harmful content and abuse? Who stands behind the creation and dissemination of fake news—foreign governments, terrorist organizations, the private sector, or someone else? What is the purpose of generating fake news—creating panic and confusion, maximizing profit in the spirit of the new pattern of surveillance capitalism, controlling certain target groups of users, or something else?

This paper is an attempt to address at least some of these issues from the Croatian perspective. For many reasons, this perspective is very peculiar. The rapid development of electronic communications over the years, underscored by the global crisis caused by the novel coronavirus (COVID-19) pandemic, has shifted citizens' professional and private lives onto virtual platforms.[8] These global trends have affected people in Croatia in a way that is not much different from other countries. However, the Croatian perspective on these issues is somewhat specific due to reasons that trace back to recent history. First and foremost, the country's transition period is ongoing. The country was born after the dissolution of former Yugoslavia, and it gained its *de facto* independence through years of bloodshed in the Homeland War in the 1990s. Human casualties, economic devastation, and direct and indirect damages to private and state property were some of the serious consequences of the armed conflict and aggression, which resulted in the cessation of one third of the territory. These regions were regained and returned to the control of the capital in 1995 and 1998 through military campaign and peace negotiation settlements. Parallel to the atrocities, 90s Croatian society underwent a shady privatization process that generated huge inequalities among citizens.[9] This negative stratification resulted in a global sense of injustice and a low level of public trust in the judiciary.[10] Un-

6 See: https://bit.ly/3tXS1XL.

7 Henson, Reyns, and Ficher, 2013, pp. 475–497.

8 Volosevici, 2020, pp. 109–116.

9 Getoš Kalac and Bezić, 2017, pp. 1091–1120.; Roksandić Vidlička, 2014, pp. 1091–1120.

10 According to the 2020 European Union (EU) Justice Scoreboard, about 60% of respondents expressed their distrust of court judges in Croatia. This is the highest percentage in the EU. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/upholding-rule-law/eu-justice-scoreboard_en; In the 2019 Global Competitiveness Report presented by the World Economic Forum (WEF), Croatia's judicial independence was ranked 126th out of 141 countries, which is the worst rating in the EU, https://www.weforum.org/reports/global-competitiveness-report-2019; However, it is not only the justice system that is perceived as problematic in Croatia. For instance, in 2020 Croatia was ranked fourth (after the Russian Federation, Turkey, and Ukraine) for the total number of violations of Article 6 of the Convention for the Protection of Human Rights and Fundamental Freedoms (fair trial), The ECHR in Facts in Figures 2020, European Court of Human Rights, February 2021).

doubtedly, this social context confused and disoriented people, most of whom use the Internet and social networks, and made them more vulnerable to victimization, abuse, and manipulation, both physically and in cyberspace.

Section 2 deals with freedom of expression and the prohibition of censorship in Croatia in the context of Internet and social media use. Section 3 provides an overview of the criminal legislation restricting free speech for the protection of some other constitutional values. Section 4 addresses legislative and institutional frameworks for regulating electronic media. The new Draft Electronic Media Act (DEMA), which introduces the concept of electronic media's responsibility for user-generated content, is analyzed in Section 5. The thematic focus of Sections 6 and 7 is on fake news in general and the legal regulation of this phenomenon in Croatia. Section 8 analyzes two issues relevant to this topic: What are the nature and scope of content provider responsibility for user-generated content? (Section 8.1.) Is there a need for *lex specialis* regulation of social networks? (Section 8.2.) Finally, Section 9 is a conclusion containing an attempt to answer the question: Does it make sense to counter fake news in a world where the truth has (almost) disappeared?

In sum, the purpose of this paper is to scrutinize and analyze regulations and procedures, identify their weak points, and offer proposals to improve dysfunctional legislation and the ineffective implementation of Internet and social network policies.

2. Freedom of expression and prohibition of censorship

Freedom of expression in Croatia is one of the greatest constitutional values. It is a political right guaranteed by Article 38 of the Constitution of the Republic of Croatia (hereinafter, the 'Constitution'). This provision guarantees freedom of opinion and expression. Freedom of expression includes, in particular, freedom of press and other means of communication, freedom of speech and public appearance, and the free establishment of all public communication institutions.[11] This constitutional provision prohibits censorship without specifically defining the term. Nevertheless, it can be concluded that the term 'censorship' refers to the performance of journalistic work or the journalistic profession because journalists' right to freedom of reporting and access to information are regulated immediately after the provision on the prohibition of censorship. In the literature, censorship refers to journalists and denotes different prohibitions on publishing certain information.

The Croatian encyclopedia defines censorship (lat. *censura*: assessment of property, assessment) as a system of administrative measures taken by the state, religion, party, and other authorities against the disclosure, reading, dissemination,

11 Ustav Republike Hrvatske, Narodne novine (Official Gazette) 56/1990, 135/1997, 113/2000, 28/2001, 76/2010, 5/2014; for an updated English version, see: https://bit.ly/3nTPCwg.

and possession, listening to, and viewing of printed and manuscript books, films, videocassettes, radio and television shows, theater plays, etc., and similar material considered undesirable and dangerous to society. There are several types of censorship: preventive censorship, self-censorship, and suspensive censorship.[12] The purpose of preventive censorship is to prevent or hinder publication by various measures such as the obligation to send manuscripts to certain bodies to verify the content, denying funding for publication costs, etc. Self-censorship is a procedure in which the authors themselves restrict their freedom of expression while facing the risk of the harmful consequences of publication. Finally, suspensive censorship manifests itself in the *ex post facto* procedures of indexation, prohibition, seizure, and other measures that restrict or prohibit the distribution of given content.[13]

In democratic societies, censorship is forbidden. It is considered unconstitutional because it restricts freedom of expression, which is among the fundamental constitutional values. However, notwithstanding the prohibition of censorship, in a democratic society based on the rule of law, certain restrictions on freedom of expression are allowed. These restrictions do not amount to censorship. Thus, the Convention for the Protection of Human Rights and Fundamental Freedoms (hereinafter, the 'Convention') states that freedom of expression may be subject to formalities, conditions, restrictions, or penalties prescribed by law, which in a democratic society are necessary in the interests of national security, territorial integrity, or public order to prevent disorder or crime, protect health or morals, protect others' reputation or rights, prevent the disclosure of confidential information, and preserve the judiciary's authority and impartiality.[14]

The European Court of Human Rights (ECHR) has established through its rich case law that any restriction on freedom of expression must be prescribed by law, necessary in a democratic society, and proportionate to the nature of the restriction.[15] Thus, the Court has repeatedly found states' margin of appreciation to be very narrow when it comes to expressions aimed at political criticism. This, in other words, means that in such situations, even statements that are offensive and shocking should be tolerated if they exercise a functional democratic right to freedom of expression.[16] This is because such freedom is a condition for the functioning of a democratic society, and without freedom of expression, the normal course of the democratic process is inconceivable. Furthermore, such a restriction must be necessary in a democratic society (necessity test), which means that certain values cannot be protected in any way other than by restricting freedom of expression. Freedom of

12 See: https://www.enciklopedija.hr/natuknica.aspx?id=11246.

13 Ibid.

14 Convention for the Protection of Human Rights and Fundamental Freedoms, https://www.echr.coe.int/documents/convention_eng.pdf; Schabas, 2016, pp. 444–482; Alexander, 2005.

15 ECHR, Observer and Guardian v. the United Kingdom, application no. 13585/88, 1990.; ECHR Prager and Oberschlick v. Austrija, application no. 15974/90, 1995.

16 ECHR, Handyside v. the United Kingdom, application no. 5493/72, 1976; Alaburić, 1996, p. 537-557; Sears, 2020, pp. 1327–1376.

expression excludes incitement to hatred and violence.[17] Hence, the Croatian Constitution prohibits any incitement to war or the use of violence, national, racial, or religious hatred, or any form of intolerance.[18]

Freedom of expression and its restrictions have been repeatedly scrutinized by the Constitutional Court of the Republic of Croatia (hereinafter, the 'Constitutional Court'), which, applying the ECHR standards, has emphasized that free speech is the one of the fundamental pillars of a democratic society. Regardless of the importance of freedom of expression, it carries with it 'duties and responsibilities' that "take on importance, among other things, when the reputation of named individuals is attacked, and the rights of others are undermined."[19] The protection afforded to journalists reporting on matters of general interest is subject to the condition that their actions are taken in good faith with the intent to provide accurate and reliable information in accordance with journalistic ethics. Therefore, in assessing reporting on matters of general interest, state bodies are limited by the democratic societal interest to enable the media to play their key role as guardians of the public interest. Freedom of expression refers not only to 'information' or 'ideas' that are favorably accepted or not considered to be offensive or those which provoke no reaction, but also to those that offend, shock, or harass. This requires pluralism, tolerance, and free mindedness, without which there is no 'democratic society.' In assessing whether there has been a violation of freedom of expression, it is necessary to consider each case in the light of all the circumstances, including the content of the allegations in question, as well as the context in which those allegations were made. In particular, it is necessary to determine whether the measures taken to restrict freedom of expression are proportionate to the legitimate aim pursued by that restriction and whether the interference is 'necessary in a democratic society.'[20]

3. Freedom of expression and its restrictions in criminal legislation

The aforementioned content-based prohibitions, which are deeply rooted in the constitutional order, are also incorporated into other regulations that protect certain constitutional values. Thus, the Criminal Code[21] contains provisions that criminalize so-called 'expressive' offenses. A long time ago, the Roman classical jurist Ulpian created the maxim that no one can be punished for his thoughts (lat. *cogitationes*

17 Cassim, 2015, pp. 303–336.
18 Supra note 11.
19 Constitutional Court, U-III-2858/2008, 22 December 2011.
20 Ibid.
21 Criminal Code, Narodne novine (Official Gazette), 125/2011, 142/2012, 56/2015, 61/2015, 101/2017, 118/2018, 126/2019.

poenam nemo patitur). However, certain manifestations of the will that are most often expressed in words (written, spoken) but can also be expressed in other ways (e.g., symbolic or real insult or *iniuria*) and which violate the rights of others, the legal order, national security, etc., are prescribed as criminal offenses and may imply criminal liability. These offenses can be divided into several categories:

- offenses of public incitement,
- offenses of breach of honor and reputation,
- offenses of breach of secrecy,
- other offenses.

For the purpose of this paper, only the first two categories will be addressed and briefly explained. There are two criminal offences in the first category: public incitement to terrorism and public incitement to violence and hatred. Public incitement to terrorism consists of the public presentation or transmission of ideas that directly or indirectly incite the commission of a criminal offense with elements of terrorism. The prescribed punishment for this crime is imprisonment from one to ten years.[22] Public incitement to violence and hatred is a crime directed against a group of people or members of that group because of their racial, religious, national, or ethnic affiliation, language, origin, skin color, gender, gender identity, disability, etc.[23] It is very similar to discrimination. In this criminal offense, *differentia specifica* is incitement to violence and hatred—an element that is not present in the criminal offense of inciting racial and other discrimination. Organizing a group (linking three or more people together) or participating in a group that conducts public incitement to violence and hatred is particularly punishable. Public approval, denial, or significant relativization (mitigation) of international crimes against a group or group member on one of the discriminatory grounds, provided that such conduct is appropriate to incite violence or hatred against individuals or the group as a whole, is also punishable as a special form of this crime.[24]

Criminal offenses of public incitement are commonly committed through a computer system. It does not matter whether certain criminalized content has been made available to the public via electronic publication, commentary, social network, etc. It must be a modality of committing an act (*modus operandi*) due to which prohibited content has become available to more people. Incriminations of public incitement should be distinguished from incitement to commit a criminal offense as a provision of the general Criminal Code. Unlike the general provision on incitement, in which the person being encouraged to commit the crime must be concretized or individualized by the instigator, public incitement is about influencing the will of unspecified individuals or groups. These individuals or groups are instrumentalized by the instigator acting as some sort of provocateur to commit the crime.

22 Ibid., art. 99.
23 Ibid., art. 325.
24 Ibid.

Insult and defamation are offenses against honor and reputation, which, based on a permissive constitutional norm, restrict freedom of expression. Insult consists of belittling another person. It is a negative value judgment or statement that is not subject to the proof of truth. The insult can be committed through words (*verbal iniuria*), signs (*symbolic iniuria*), or deeds (*real iniuria*). It is an aggravated form of the offense if someone insults another so that the insult becomes known to a larger number of people. One way is to commit such an act via a computer system or network. The penalty for insult is only monetary, and the procedure is initiated by private lawsuit.[25]

Unlike insult, defamation must be about creating or disseminating a false factual statement or content with respect to which truthfulness can be equally established for all persons (e.g., claiming that someone was previously convicted). Defamation is a criminal offense in which the burden of proof is shifted to the defendant. This is an understandable exception to the principle of *actore non probante reus absolvitur* given that the plaintiff cannot be expected to prove a negative fact. As all people are considered good (*quisquis presumitur bonus*), those who claim otherwise must prove it. Like insult, defamation is also punishable only by a fine, and criminal proceedings are initiated by private lawsuit. The commission of the offense via a computer system or network is considered an aggravated form for which a heavier penalty (also a fine) is prescribed.[26]

Apart from the aforementioned criminal offenses, some other laws have imposed permissible restrictions on freedom of expression. The regulation relevant to this analysis is the Law on Misdemeanors against Public Order and Peace, which prescribes the criminality of certain types of behavior in a public place, including the dissemination of false news. Section 7 analyzes this offense, together with some other criminal offenses (e.g., false alert).

4. The legislative and institutional framework of electronic media

In the context of freedom of expression on the Internet, regulations and practices relating to electronic publications are particularly relevant. The increase in the number of electronic publications in the world in the last few years is significant, and such a trend is noticeable in Croatia, too. DEMA reveals that the number of electronic publication providers has increased from 276 to 336 in a short period of time.[27] According to the Electronic Media Act (EMA), electronic publications are

25 Supra note 21, art. 147.
26 Supra note 21., art. 149.
27 Draft Electronic Media Act. Available at: https://bit.ly/2XvznKO.

a type of electronic media, in addition to audiovisual and radio programs. They are published daily or weekly for the purpose of public information and education. A media service provider is a natural or legal person who has editorial responsibility for selecting audio and audiovisual content and audio and audiovisual media services and determining how these are organized.[28]

Based on the constitutional norm, EMA guarantees freedom of expression and full programmatic freedom for electronic media, and no legal provision can be interpreted as granting the right to censor or restrict the right to freedom of speech and expression. However, the law prohibits media services from threatening the constitutional order and national security, inciting hatred and discrimination, expressing the ideas of fascist, nationalist, communist, and other totalitarian regimes, disclosing private family life pertaining to children, etc.[29]

The central regulatory body for the implementation of the EMA is the Council for Electronic Media (hereinafter, the 'Council'). The Council is a part of the Agency for Electronic Media (hereinafter, the 'Agency'). It manages the Agency and performs the regulatory body's tasks in the field of electronic media. The president and members (seven in total) perform their duties as full-time employees. The transparency of the Council's work is ensured through the annual submission of reports to the Croatian Parliament. The Council's president and members are appointed and dismissed by the Croatian Parliament on the proposal of the Government of the Republic of Croatia. In proposing Council members, the Croatian government announces a public call. Members' term of office is five years, and they are eligible for reappointment. The formal criterion that should be met for appointment concerns citizenship (Croatian). Substantive criteria concern professional knowledge, abilities, experience in radio, television, or publishing, or cultural or similar activities. The status of a state official, an official in the executive or judicial branch, and a political party official is not compatible with Council membership.[30] The Council's responsibilities include, *inter alia*, granting concessions to the electronic media, decision making in cases of the revocation of concessions and permits, issuing warnings in cases of non-compliance with the provisions of the EMA and the bylaws, and/or reporting cases to other competent authorities (e.g., misdemeanor courts). The Council is also tasked with encouraging media literacy, organizing public consultations and professional gatherings, and conducting research related to the functioning of electronic media. The Council's decisions cannot be appealed, but there is the possibility of initiating an administrative dispute before the competent administrative court.[31]

The annual reports submitted to Parliament for adoption are the best source of information about the Council's activities. However, media coverage of their decision making is sporadic and on an ad hoc basis, and the Council's website is not

28 Electronic Media Act, Narodne novine (Official Gazette) 153/2009, 84/2011, 94/2013, 136/2013.
29 Ibid.
30 Ibid.
31 Ibid.

regularly updated nor does it contain current information about their activities.[32] This is certainly something to be criticized because the mechanism in charge of evaluating the media's transparency procedures should be much more transparent than it has been so far. As to the measures the Council applies, the 2019 report states that based on citizens' complaints or upon *proprio motu* supervision, the Council issued three measures pertaining to the violation of the provisions on advertising and sponsorship, 14 measures concerning the violation of concession obligations and legal program minimums, 15 measures related to the violation of the legal provisions for the protection of minors, one measure regarding spreading and inciting hatred or discrimination, and one measure addressing an uncategorized violation of the law. All 34 measures, except one pertaining to the permanent revocation of the concession, represent warnings/admonitions. Two cases were forwarded for further proceedings to the Croatian Journalists' Association (HND) and the municipal state attorney's office in Varaždin.[33]

In an illustrative warning case pertaining to a measure the Council applied, a decision was issued against a publisher (a television channel) in 2019. The Council found that there had been spread and incitement to hatred and discrimination in response to a viewer's complaint about the display of a flag bearing fascist (Ustasha) symbols (the letter 'U') attached to a central news broadcaster on January 21, 2019. The Council issued another warning in a different case that was ruled upon in 2019. The case concerned a publisher (a television channel) that violated Article 12, paragraphs 1 and 2 of the Ordinance on the Protection of Minors in Electronic Media. The disputed show featured an Arab physician and exorcist who spoke about healing diseases caused by spells, magic, and demons. The show aired at 6.15 pm in violation of the provision that programs depicting gambling, fortune telling, alternative healing methods, and other similar issues/services that are not scientifically based cannot be broadcasted before 11 pm, at which time it is mandatory that they bear an appropriate 'graphic' label/notice.[34]

In the past, the appointment of Council members was controversial and fueled widespread public debate. For example, in June 2019, the HND opposed the possibility of appointing a member belonging to the Croatian Journalists' and Publicists' Association (HNIP)[35] due to Association funding of those authors, *inter alia*, whose shows had been banned by the previous Council due to hate speech. Those same authors and their banned shows were the reason the former Council president resigned in March 2016. Specifically, at the end of January 2016, the Council decided to stop broadcasting on local television for three days. The decision was made due to

32 See: https://www.aem.hr/vijece/.

33 Izvješće o radu Vijeća za elektroničke medije i Agencije za elektroničke medije za 2019. godinu. Available at: https://bit.ly/39tp8Ju.

34 Ibid.

35 See: https://bit.ly/3EAUw7t.

hate speech espoused by the journalist M. J., host and editor of the show 'M.T.,' who closed the show with the following words:

> Two Chetniks who were priests of the Serbian Orthodox Church were canonized in 2005. According to witnesses, they bled their hands. We do not know whether the Serbian Orthodox Church will continue to do so. Therefore, the people of Zagreb who walk through Flower square should be careful, especially mothers with small children. Be careful when passing by the Church so that no one runs out with a knife and performs their bloody Chetnik feast.[36]

The Council's decision triggered demonstrations by a number of protesters (including Homeland War veterans) who claimed the decision had reintroduced the communist regime's verbal delict in Croatia. The Council president was presented with a Chetnik and partisan hat as provocation.[37] She construed this act as pressure and decided to resign.[38]

The other controversial case upon which the Council ruled concerned a web portal. Following the public display of the body of the deceased saint L. M. in Zagreb, the journalist H.M. wrote an article titled "Dead in live: Catholic necrophiliac orgies are the craziest show on HRT (Croatian Television)."[39] The case was reported to the Council, who found that while the speech was offensive to Catholics, it did not amount to hate speech as prohibited under international and domestic standards. The Council quoted some ECHR standards concerning the distinctions between hate speech and other offensive and shocking statements that could appall certain groups and individuals but are protected under the function of free speech. In other words, the Council established that the alleged article "has the function of an exhaust or safety valve in a democratic society and does not deserve sanction." The Council also pointed out that the concept of an insult does not apply to the religious community and faith in general: "To be an insult, the expression must be directed at any particular person to whom such expression could harm honor and reputation, and this text ridicules the Catholic Church and the faithful in general."[40]

This reasoning is flawed and shows substantive inconsistency in the Council's reasoning. First, it is striking that compared to the positive decision in the aforementioned television-related case where there was found to be a violation that was sanctioned in just a few lines without any solid reasoning whatsoever, the Council's decision regarding the web portal involved extensive reasoning in support of a negative decision and the finding that there was no violation.[41] This is problematic because the facts of these two cases are not as different as they may initially seem. They

36 See: https://bit.ly/3lD6Zig.
37 See: https://bit.ly/3Cx7KQS.
38 See: https://bit.ly/3tY2i6e.
39 See: https://bit.ly/3zvraDH.
40 Ibid.
41 Ibid.

both concern the general insult and ridicule of a religious community and a church as an institution. Therefore, it does not seem plausible for the Council to find a tendency toward hate speech only in the first case rather than in both. Furthermore, the Council completely overlooked another important aspect of the Convention: the rights and freedoms under Article 9. The ECHR expressed that some criticism of religion, in general, should be considered as a vehicle fostering public debate about issues of common societal interest.[42] However, extreme satirical (as well as other artistic and non-artistic) expressions made with the sole purpose of provoking and insulting others' religious feelings lack that function (i.e., speech that is 'gratuitously offensive'), and states must have a means to restrict their reliance on the margin of appreciation doctrine. Ostensible artistic or quasi-artistic nature of expression must not be used as a *carte blanche* to violate others' rights.[43] In its further jurisprudence, the Council should take this into due consideration. Otherwise, its decision making will remain subject to further criticism due to inconsistency and the upholding of double standards.

In cases of suspected violations, the Council can initiate proceedings on its own, without a formal report/claim. However, there are no clear and transparent criteria for establishing a threshold for conducting such *proprio motu* proceedings. A good example is a recent case involving a web portal founder. In his Facebook status, he called for the demolition and burning of churches and government buildings due to his dissatisfaction with COVID-19 restrictions and alleged cases of non-observance. He wrote, *inter alia*, the following:

> If there were a world championship in accepting humiliation, Croats would win first place. Every unbroken window on the government and church buildings, every building of theirs that is not on fire even after this, every head that remains on their shoulders after all while you are free to rope, is proof that you can make as many jerks as you want from Croats. Whatever you take away from them, you can do even more.[44]

A nongovernmental organization (NGO) reported the posting of this statement to the public prosecutor's office under Article 325 of the Criminal Code (public incitement to hatred and violence). Nevertheless, it is unclear whether the Council initiated the proceedings *ex officio* for the suspected violation of the EMA as gleaned from public sources. An argument against their jurisdiction in the case might be that this text was not published in an electronic publication but on the site founder's personal Facebook account. However, such an argument would not be plausible because there is an obvious link between the site founder's expressions on his social network account and the commercial Internet news portal's economic interest. Such cases

42 ECHR Otto Preminger Institute v. Austria, application no. 13470/87, 1994.
43 Derenčinović, 2018, pp. 194–212.
44 See: https://bit.ly/3lOwZrb.

should not be easily dismissed based on the lack of jurisdiction argument because in terms of the prohibited content, there is no *de facto* difference between the official social network account on the portal and its founder's account and the person who pulls the strategical and tactical strings behind the scenes.

In summary, the current legal and institutional framework concerning Croatia's electronic media seems to be quite sound. Unfortunately, this is only theoretically true . In reality, there have been numerous implementation problems, along with apparent politicization, ideological clashes, and double standards. Much more should be done regarding the transparency of the Council's work (ad hoc sensational media coverage and annual reports submitted to Parliament are insufficient), the quality of its work through raising public awareness, the training and education of Council members (on media law and standards developed in ECHR case law), better coordination with other competent authorities to address cases of alleged hate speech and content that may be harmful to minors, etc.

5. New Draft Electronic Media Act

International standards governing the conditions under which content providers establish and manage electronic publications have been enshrined in European law. Among these are, first and foremost, Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation, or administrative action in the member states concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities.[45] The 2013 directive was amended to fill the gaps regarding the legal regulation of video-sharing platforms on which users generate their own content, as the content may be contrary to the interests, needs, and protection of children and young people. Therefore, the new directive has improved the protection of minors from harmful content and the protection of all users from hatred, violence, and incitement to terrorism. An additional purpose is to increase users' media literacy. Media literacy refers to the skills, knowledge, and understanding that enable citizens to use media effectively and safely.[46]

To harmonize national legislation with the amended directive, the Croatian Parliament is passing a new EMA. DEMA's explanatory memorandum states, *inter alia*, that it:

45 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation, or administrative action in the Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities. PE/33/2018/REV/1. OJ L 303, 28.11.2018, p. 69–92. See: https://eur-lex.europa.eu/eli/dir/2018/1808/oj.

46 Ibid.

> stipulates that audiovisual advertising must be easily recognizable as such and must not use subconscious techniques; question human dignity; include or promote discrimination; encourage behavior that is detrimental to health or safety; encourage behavior that is highly detrimental to the environment. Furthermore, audiovisual media services must not contain incitement to violence or hatred directed against groups or members of a group based on discrimination based on sex, race, color, ethnic or social origin, genetic characteristics, language, religion or belief, political or any other opinion, belonging to a national minority, property, birth, disability, age, sexual orientation or citizenship, as well as content that provokes the commission of a terrorist offense. As a novelty, this Act also introduces providers of video-sharing platforms who must take appropriate measures to protect minors from content that could affect their physical, mental, or moral development and the general public, from incitement to violence or hatred or public provocation to commit a terrorist offense.[47]

The Draft Act is still undergoing the parliamentary procedure, and it is expected to be sent to its final reading in the near future.

DEMA has also undergone the (online) public consultation procedure, through which several major issues have been identified First, there is concern that some of the provisions are vague. For example, the obligation to respect human dignity, noncompliance with which results in sanctions, raises many questions as to what belongs under the heading of 'human dignity.' Some commentators have pointed out that human dignity is not fully protected by media legislation anywhere in the world and that in some situations, it is permissible to question certain categories' dignity (e.g., politicians') in the interest of public debate. Hence, there have been concerns that such a vague provision could negatively affect freedom of expression and lead to self-censorship.[48] In principle, this is true. However, there was little maneuvering room for drafting DEMA given that the directive itself stipulates that audiovisual commercial communications must not jeopardize respect for human dignity. In any case, this provision would, in practice, require the delicate balancing of competing interests. Another argument related to DEMA is that it significantly intensifies repression, which will, in challenging the conditions of the media's functioning in Croatia, shut down many publishers in the electronic environment.[49] The number of violations that DEMA envisages is higher than that of EMA. Nevertheless, the maximum fine is still HRK 1,000,000.00 per legal entity, and this has not been changed. For regulator standards regarding the imposition of measures, see supra Section 4.

DEMA's most important novelty is Article 93, paragraph 3, which states that the provider of an electronic publication is responsible for all the publication's content, including that generated by users.[50] This provision has caused the most doubts and

47 Supra note 27.
48 See: https://bit.ly/3At1RTT.
49 See: https://bit.ly/39nZbLt.
50 Supra note 27., article 93. par. 3.

spawned a host of diverse interpretations. There have been concerns that extending service providers' responsibilities to user-generated content will call their normal functioning into question. Specifically, for web portals with many followers, it will be difficult to expect editors to peruse all the comments and, if necessary, filter those with inappropriate or banned content. Several factual and legal questions also arise here: According to the draft provision, will service providers be obliged to verify certain content's veracity? Will they have to determine whether certain content is, for example, incitement to violence or hatred?

The legal basis for the attribution of responsibility for others' acts, over which the content provider has limited supervision possibility, is disputable. This solution is also questionable when it comes to the personal culpability standard. The mentioned criminal offenses are, without exception, punishable only when they are committed with intent, which means that the perpetrator either seeks to achieve a prohibited consequence (e.g., incitement to discrimination) or at least agrees to it (*dolus eventualis*—it is immaterial whether the consequence occurs). On the other hand, a content provider's liability may be based on an omission in which there are no elements of intent but where they may be negligence. This raises several doubts about how to assign responsibility to the publisher for user-generated content while maintaining the principle of guilt intact. There have been concerns that this could result in terminating the comments feature online. Given that many readers are attracted by the ability to publish comments, some electronic media fear that abolishing commenting will render their portals less interesting to the public.

On the other hand, there is no doubt that electronic publications generate revenue on the market not only through the content they offer to the public, but also through user-generated content in the form of comments. This is the logic underlying paid advertisement: the more clicks certain links receive, including comments, the higher the content provider's advertising revenue. If revenue is generated in this manner, i.e., the monetization of the activities on a given portal, then the same media should be considered responsible for the anonymous comments that help them generate profits. However, by introducing registration as a prerequisite to using the commenting feature on articles and other content, publishers would be exempted from liability, and possible criminal or other proceedings (e.g., civil proceedings for personal rights violations) could be initiated against persons listed as the authors of given statements. The purpose of this new legislative approach is to limit anonymous comments and introduce a system of responsibility for the spoken word in those cases where it is contrary to the principle of freedom of expression or when it serves to spread hatred and violence, discrimination, the abuse of children and youth, incitement to terrorism, etc.

This is a very delicate issue involving competing rights and interests, and the interpretation of this provision will depend not only on the circumstances of the given case, but also on the standards pertaining to electronic media's responsibility for user-generated content established in ECHR jurisprudence (see the discussion and conclusion on this issue infra in Section 8.1.).

6. The concept of fake news

The fake news phenomenon has been described in the literature as "spreading outrageous distorted information to discredit opposition or create divisiveness between opposing groups."[51] According to another definition, fake news is "false, often sensational, information disseminated under the guise of news reporting."[52] The European Union (EU) defines disinformation as "verifiable false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public, and may cause public harm."[53] For some commentators, the term is defined through the consequences it produces as the destabilization of the category of truth in a democracy for geopolitical gain.[54] According to the some consequentialists, the fake news phenomenon creates an environment where "emotion triumphs over reason, computational propaganda over common sense, or sheer power over knowledge."[55]

There have been some suggestions in the literature to resolve the conceptual confusion concerning the terminology related to fake news. In this regard, three similar types of information can be identified. The first is misinformation, which is false, although there was no intention behind its creation. Unlike misinformation, another type—disinformation—is characterized by its creator's intention to deliberately harm others. Finally, there is malinformation, which is neither created nor fabricated. It exists in social reality (hate speech), but similar to disinformation, it is intended to harm others.[56]

Fake news is not a 21st century invention. Its origin significantly preceded the Internet and social networks. Often deliberately placed, fake news has influenced the course of historical events. An interesting example from recent history is taken from historiography and known as the 'Ems telegram.' On the brink of the Franco-Prussian War, a meeting took place in 1870 between Prussian Emperor Wilhelm I and French Ambassador Vincent Benedetti. On that occasion, Benedetti kindly asked the Prussian emperor to relinquish his claim to the Spanish throne to his family members, and Wilhelm I, even more kindly, refused. Otto von Bismarck was informed of the brief courteous meeting via a telegram that described the whole event as an incident. By revising the telegram's original text and giving it a more conflicting tone, Bismarck provoked France into declaring war on Prussia. The fake

51 Nielsen, 2020, cited in Dalkir and Katz, 2020, pp. 238–257.
52 See: https://www.collinsdictionary.com/dictionary/english/fake-news.
53 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions tackling online disinformation: a European approach. com/2018/236 final. See: https://bit.ly/3tXUKAt.
54 Mueller, 2019.
55 Peters, Rider, Hyvönen and Besley (eds.), 2018.
56 Wardle and Derakhshan, 2017.

news (or distorted truth with many layers of lies) became the *casus belli* that affected European history in the late 19th century.[57]

Another example from more recent history shows how the creation and dissemination of fake news has taken on the proportions of a mass phenomenon. This is clearly demonstrated by totalitarian regimes' propaganda. These are industries of lies calculated to discredit the opposition, i.e., groups perceived as potential points of resistance. Nazi propaganda utilized elaborate methods known as the phenomenon of the 'lying media' (Lügenpresse).[58] Similarly, communists fabricated 'news' about the ideal life behind the Iron Curtain. These were shallow stories about societies free of crime and social pathology. The reality was societies that celebrate and promote equality and a (distorted) version of human rights.

7. Legal regulation of fake news in Croatia

As espoused in Section 6, the creation and dissemination of fake news is not a new phenomenon; however, in the era of the technological revolution and global digitalization, it has taken on massive proportions. Thanks to social networks, it has never been easier or faster to share content, including fake news. An interesting source of information about fake news in Europe is the Eurobarometer research conducted in 2018 among 26,576 respondents from 28 EU member states. They were interviewed about their trust in media in general and about electronic media in particular. The results show that traditional information sources are the most trusted (television, radio, and print media), while social networks and messaging applications are much less trusted (26%).[59]

Like their counterparts in other countries, respondents from Croatia mostly trust traditional media (for instance, 65% of them trust television, which is close to the EU average of 66%). When it comes to social media and messaging applications, 36% of respondents in Croatia do not trust them. The European average for distrust is 54%. Only respondents from Estonia, Lithuania, and Romania have more trust in news and information accessed through online social networks and messaging applications than those in Croatia. It is interesting that despite the higher level of trust in digital media in Croatia compared to most other European countries, almost 76% of the respondents come across news or information believed to be fake at least once per week. This is above the European average (68%). According to the respondents, the situation seems to be worse only in Spain (78%) and Hungary (77%).[60]

57 See: https://www.britannica.com/event/Ems-telegram.
58 Supra note 51, p. 239.
59 See: https://bit.ly/39q4F8I.
60 Ibid.

At the same time, the respondents in Croatia seem to be among the most (self-) confident in terms of their perception of their ability to identify fake news (82% in comparison to the EU average of 71% and exceeded only by respondents in Denmark [87%] and Ireland [84%], who are more confident in this regard). Most of Croatian respondents believe that the fake information phenomenon is a problem (86%, compared to the EU average of 85%) and that it poses a threat to democracy (80% compared to 83% in the general sample).[61] The total number of respondents in Croatia was 1,005, and there were some differences concerning regional distribution, age, and place of residence (urban vs. rural areas). An interesting but expected finding is that trust in the media decreases with age. Older persons have less trust in electronic media specifically, while television and radio enjoy a high level of trust irrespective of respondents' age.[62]

The phenomenon of fake news on the Internet and in social networks in Croatia flourished after the outbreak of the COVID-19 pandemic. The Council issued the following warning:

> All audiovisual media services are banned, including those via the Internet that publish or spread misinformation, especially those related to public health issues. Disclosure or dissemination of misinformation causes concern, the spread of fear and panic among the population and leads to even more severe consequences than those we face.[63]

The state attorney office also released a statement concerning the placement of false information in the aftermath of the pandemic outbreak. The purpose was to instruct all county and municipal state attorney offices "to act thoroughly and immediately in accordance with the provisions of Article 38, paragraphs 1 and 2 of the Criminal Procedure Code and Article 35 of the State Attorney's Office to detect and prosecute perpetrators."[64]

The following are some of the competent authorities' reactions to fake news that appeared on social networks concerning public health issues. In one instance, some 'well informed' citizens were offering 'confidential' information to the public via social media about when complete quarantine would occur, allowing persons to leave the house once per week. Others 'reliably knew' which stores/hospitals housed COVID-19 infected people, and they felt that this knowledge exempted them from self-isolation. Other cases concern a woman who falsely introduced herself as a doctor and published 'real' news about the spread of the virus, a night watchman who posted photos from previous gatherings on social media during his shift and claimed that they were happening in the present in violation of measures implemented to prevent

61 Ibid.
62 Biloš, 2020, p. 166–185.
63 See: https://bit.ly/39lOQje.
64 See: https://bit.ly/2XA62Pe.

the spread of the infection, a woman who falsely announced her own COVID-19 infection on her social media site, etc. In these and similar cases (32 in total), the police filed misdemeanor reports with the competent misdemeanor courts.[65]

Following the series of severe earthquakes in Zagreb and surrounding areas from March 2020 onward, fake news concerning future earthquake predictions began to spread via social media. Although official geologists and other competent authorities explained that earthquakes cannot be accurately predicted, the fake news continued, creating confusion and panic in the citizenry. The police opened several cases, including one concerning a suspect who was persistently publishing false and disturbing news/content about earthquakes through a social network and a network for publishing and exchanging video clips.[66] Fake news was also disseminated through social networks in the immediate aftermath of the strong earthquake that hit central Croatia (Petrinja and surroundings) on 29 December 2020. This was in the form of a photo that showed large-scale damage and destruction in Petrinja, a city about 60 km southeast of Zagreb, the earthquake's epicenter. However, the photo was false because the city depicted was not Petrinja but Amatrice, a town in Italy that was also hit by a devastating earthquake in 2016.[67]

The recent increase in the incidence of fake news on the Internet and social networks does not mean that this phenomenon has not been present in Croatia for quite some time. During the 2019 European parliamentary elections, two cases of alleged fake news perpetrated by the candidates were reported to the Ethics Committee of the Croatian Parliament. The first concerned a candidate who published on her official website that an association active in promoting patients' rights was "against vaccination and supports anti-vaxxers." The Ethics Committee compared this statement with the association's official webpage (also a signatory to the Declaration on Compulsory Vaccination of the World Health Organization) and concluded that the statement the candidate published was untrue. Hence, the Committee found that there had been a violation of the Code of Ethics in the Elections.[68]

In another case, the Committee concluded that there had been no violation. This case was about a candidate's allegedly false paid advertisements on social networks. The advertisements claimed that according to the polls, the list would win three seats in the European Parliament. The Committee established that in legal terms, such advertising could not be considered false. The number of seats that candidates will win is an uncertain future fact, and candidates are free to make estimates, including their own predictions about the number of seats. The Committee also pointed out that it has no instruments to determine whether a particular advertisement is true or false.[69]

65 See: https://bit.ly/3tZq0iP.
66 See: https://bit.ly/3AxElFt.
67 See: https://bit.ly/3hP2NdZ.
68 See: https://www.izbori.hr/site/UserDocsImages/1975.
69 Ibid.

Creating and disseminating fake news is prohibited under Article 13 of the Act on Misdemeanors Against Public Order and Peace.[70] Any person who invents or spreads false news that disturbs citizens' peace and tranquility can be punished with a fine. This legislation dates back to 1977, so the fine is still prescribed in the defunct currency German marks (due to the high inflation at that time in the former Yugoslavia). This means that whenever the sanction is imposed, the court calculates the corresponding amount in the domestic currency. In addition to the misdemeanor against public order and peace, some fake news-related behavior could also be qualified as the criminal offense prescribed in Article 316 of the Criminal Code, namely false alert.[71] The perpetrator is whoever falsely informs the police or another public service that ensures order or provides assistance about an event that requires urgent action under that service. False alert can be punished with imprisonment of up to three years.[72] The most common *modus operandi* for a false alert in previous years was a false report about planted explosive devices, forcing the police and other competent authorities to evacuate buildings to prevent loss of lives and property damage.[73] The *ratio legis* for this criminal offense consists of the high expenses the police incur when conducting their interventions without valid reason, as well as their diversion from other law enforcement activities. Given the pandemic crisis, some commentators have suggested that causing public panic for no reason should be a consideration for increasing the penalty in future legislative amendments.[74] However, there have been no initiatives to reintroduce the creation and dissemination of fake news as a separate criminal offense. Nevertheless, under certain circumstances, the spreading of fake news could be legally qualified under Article 316 of the Criminal Code in cases where the police or other competent authorities were activated to conduct an inquiry or investigation.

Over the last several years, there has been an increase in various projects in Croatia dealing with preventive aspects of the fake news phenomenon. October 2020 saw the launch of the website 'Museum of Fake News.' It is envisaged as a repository of documents, essays, and other materials concerning the topic, and it also offers useful tools for self-prevention (also self-protection) and the promotion of media literacy.[75] According to the initiative's authors, the purpose of the website is to make citizens aware of the prevalence of fake news, raise media and information literacy levels, educate citizens about how to recognize fake news, etc.[76] Another interesting

70 Law on Misdemeanours against Public Order and Peace, Narodne novine (Official Gazette) 47/1990, 55/1991, 29/1994.
71 Supra note 21., article 316.
72 Ibid.
73 See: https://bit.ly/3lJ92kQ.
74 Moslavac, B., Lažna uzbuna i lažne vijesti, https://www.iusinfo.hr/strucni-clanci/lazna-uzbuna-i-lazne-vijesti.
75 See: https://www.croatiaweek.com/croatia-to-get-museum-of-fake-news/.
76 See: https://mlv.hr/o-nama/.

website is 'Media literacy,'[77] founded by the Agency for Electronic Media and the United Nations International Children Emergency Fund (UNICEF) in partnership with several interlocutors from academia and NGOs. The website is mostly designed for younger people, including students, with a focus on those who are the most vulnerable to harmful content and therefore the most in need of media literacy. Major thematic focuses include the prevention of disinformation, safety on the Internet, children and the media, the media and violence, and the media and stereotypes.[78]

Having been a member state since 2013, Croatia has joined the EU initiatives. Within the EU, various types of information disorder have been taken seriously. An action plan against disinformation was adopted in an effort to ensure that the 2019 European parliamentary campaigns would be free of disinformation and fake news.[79] The action plan clarifies that fake news and disinformation campaigns are part of hybrid warfare.[80] Those behind such campaigns include some foreign governments and non-state actors. The latter are mostly involved in spreading vaccination-related false news . The measures envisaged in the action plan are divided into four categories: improving the capabilities of Union institutions to detect, analyze, and expose disinformation; strengthening coordinated and joint responses to disinformation; mobilizing the private sector to tackle disinformation; and raising awareness and improving societal resilience. The dissemination of disinformation and growing populism were highlighted as thematic areas of interest for the EU in the Priorities of the Croatian Presidency of the Council of the European Union (1 January–30 June 2020).[81] Various organizations sent their comments to the Priorities. For instance, the Commission of the Bishops Conferences of the European Union (COMECE) indicated that:

> [A] rights-based approach should be promoted in any EU initiative to counter disinformation…definitions must be sharp and prevent unwanted effects on free expression and democratic debates…and self-regulation can be effective only as a complementary element…the key role must remain with the justice system.[82]

The explanatory opinion requested by the Croatian presidency titled "The effects of campaigns on participation in political decision-making," adopted in June 2020,

77 See: https://www.medijskapismenost.hr.

78 Ibid.

79 Joint communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Action Plan against Disinformation, JOIN/2018/36 final. Available at: https://bit.ly/3ED6tcB.

80 Joint communication to the European Parliament and the Council, Joint Framework on countering hybrid threats: a European Union response, JOIN/2016/018 final. Available at: https://bit.ly/3hUHrfz.

81 Priorities of the Croatian Presidency of the Council of the European Union. Available at: https://bit.ly/3hV4OFQ.

82 Contribution of COMECE and CEC to Croatia's EU Council Presidency Programme "A strong Europe in a world of challenges," January 2020, p. 8.

expressed support for the EU's efforts to counter disinformation, both external and domestic. The Commission was urged to:

> ensure full compliance and follow-up regulatory action in respect of the Code of Practice on Disinformation, further development of the recently established 'rapid alert system' and STRATCOM's intelligence units, and an expansion of the European External Action Service's action against disinformation.[83]

Croatia is among the 15 member states that signed a letter of concern regarding the spread of fake news and conspiracy theories about 5G technology in Europe. The letter sent to the Commission highlights that the EU needs "to come up with a strategy to counter disinformation about 5G technology or risk false claims derailing its economic recovery and digital goals."[84]

One of the issues relevant in the context of Croatia and other countries concerns the advantages and disadvantages of adopting a special law that would regulate unacceptable behavior (including the dissemination of fake news) on social networks. Existing legislation in Croatia is limited to the EMA, which does not regulate the rights and obligations regarding communication on social networks. The adoption of a law that would regulate social networks was announced in 2018, and, in 2019, it was included in the legislative activities plan. Some commentators supported the adoption of such a law

which would regulate the obligations of social networks to monitor and act on user reports when there is a suspicion that the user's statements committed one of the listed criminal offenses considered a priority for the Croatian legislator.[85]

Those who advocate adopting a special law believe that the Croatian law should be drafted on the model of the German Law on Law Enforcement on Social Networks (*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*, NetzDG).[86] In short, proponents of the German model stress that the new law should include social network service providers' obligation to filter content related to public incitement to violence and hatred, child pornography, and public incitement to terrorism.

On the other hand, due to some negative side effects during the implementation of NetzDG and the potential negative consequences to the constitutionally protected freedom of expression due to the risk of over-filtering, there is a majority in favor of maintaining the status quo. Thus, the 2019 Ombudsman's Report states that the approach of monitoring relevant policies at the EU level should be supported to avoid the "multiplication of national regulations governing hate speech on the Internet and fragmented regulation, which leads to unequal protection of citizens in the EU." The

83 The effects of campaigns on participation in political decision-making, rapporteur: Marina Skrabalo, SOC/630-EESC-2019, https://bit.ly/3lIS4TM.

84 See: https://www.reuters.com/article/eu-telecoms-int-idUSKBN2740Q3.

85 Roksandić and Mamić, 2018, pp. 329–357.

86 See: https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html.

report also highlights the reservation regarding adopting a special law concerning "remarks made in countries that have adopted national laws on unacceptable online behavior, but also the fear that regulation based on the German model could result in (self) censorship."[87] Many information technology experts and NGO members reacted similarly to the announcement of the adoption of a special law. One commentator pointed out that the law cannot prevent the spread of hatred and that such initiatives are proposed by people who are not sufficiently familiar with the functioning of the Internet and social networks.[88] Critics are unanimous in claiming that there is no need to enact a new piece of legislation given that prohibited conduct is clearly defined through the Criminal Code, EMA, and editorial responsibility.[89] Furthermore, they suggest that "Hate speech cannot be solved partially, only on social networks, but holistic and complementary solutions should be considered, which include civic education, media literacy and efficient and fast sanctioning of the most severe forms of hate speech."[90]

There have been no significant legislative activities since the initial announcement of the drafting of the special law, which might suggest that this idea has been abandoned, at least for the time being. Given that this is a crucial issue in the context of freedom of expression on the Internet, analysis of this issue is left for Section 8.2..

8. Discussion

8.1. What are the nature and scope of content provider responsibility for user-generated content?

There have been three important cases decided by the ECHR that concern Internet content providers' intermediary liability: Delfi v. Estonia (2013, 2015), MTE and Index.hu v. Hungary (2016), and Pihl v. Sweden (2017). In Delfi v. Estonia, the ECHR found that the state had not violated Article 10 (right to freedom of expression) when it established the media's or the publisher's responsibility for reader comments containing hate speech toward a transport company (SLK) and a member of its supervisory board.[91] Delphi is an online news portal that publishes more than 300 news items daily. It allows readers to comment and automatically posts these comments immediately after they are written, without additional portal-supervised

87 See: https://www.ombudsman.hr/hr/izrazavanje-u-javnom-prostoru/.
88 See: https://bit.ly/2XC3f8t.
89 See: https://bit.ly/39lPe1a.
90 See: https://bit.ly/3hROfur.
91 ECHR Delphi v. Estonia, application no. 64569/09, 2015.

editing or deletion. The site argued that readers who leave comments are personally responsible for the content. Delphi's site has a feature that allows other readers to label comment content as offensive or inciting hatred; flagged content is deleted. There is also a mechanism that automatically detects and deletes obscenities. In the present case, at the beginning of 2006, 185 comments on an article about SLK were published, 20 of which contained personal threats and offensive language against L. L's lawyers requested the removal of those comments, accompanied by a monetary claim of EUR 32,000.00 for non-pecuniary compensation. The disputed content was removed six weeks after publication, but the portal refused to pay the compensation.

In domestic proceedings, the Internet portal was declared liable under the provisions of the Civil Obligations Act for publishing offensive value judgments insulting another person's honor and failing to remove such content on its own initiative. The ECHR found that the impugned comments constituted hate speech, which do not enjoy protection under Article 10 of the Convention. Moreover, the Delphi portal is a professional Internet news portal that, for commercial reasons, tries to attract as many comments as possible, even on neutral topics. Given the portal's obvious economic interest regarding comments, the ECHR concluded that the portal did not function merely as a passive technical service provider. The portal's filtering measures clearly did not offer sufficient protection against speech that openly spread hatred toward L. The ECHR found the same concerning the prolonged delay in removing the disputed comments and indicated that eventual removal was on someone else's initiative. Ultimately, the ECHR found (with two separate and dissenting opinions) that there had been no violation of Article 10 of the Convention in the present case.[92]

The ECHR reached the opposite conclusion in MTE and Index.hu v. Hungary in 2016.[93] Plaintiffs were a self-regulatory body of Internet content providers and a major news portal. In this case, the ECHR found that the national courts violated the publisher's freedom of expression when they found them responsible for readers' comments on an article about a real estate company's allegedly ethically questionable advertising practices. After establishing their responsibility in civil proceedings before national judicial authorities in which the plaintiffs were awarded monetary compensation (a constitutional complaint was filed against these judgments, but was eventually rejected as unfounded), MTE and Index.hu addressed the ECHR with the argument that the state had disproportionately restricted their rights under Article 10 of the Convention. The ECHR preliminarily reiterated the standards established in its case law that Internet portals that publish news have certain rights and obligations that differ to some extent from traditional publishers', especially when it comes to content generated by third parties (commentators). However, the ECHR found

92 Ibid.

93 ECHR Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, application no. 22947/13, 2016.

an essential distinction related to the Delphi case in terms of the disputed comments' content, which, although vulgar and potentially subjectively offensive, does not constitute hate speech or incitement to violence. Unlike the Delphi case, the first applicant has no clear economic interest in monetizing its web activity through user-generated content. Finally, the comments were removed promptly, without major damage to the allegedly injured parties' protected rights. In conclusion, the ECHR found that the domestic courts had failed to conduct a proportionality test between the conflicting rights and interests. Accordingly, this amounted to a violation of Article 10 of the Convention.

Some commentators on this judgment pointed out that the ECHR had corrected its views as expressed in the Delphi case and thus protected freedom of expression in favor of electronic publication service providers. However, as Judge Kuris rightly pointed out in his dissenting opinion, this judgment in no way derogates from the standards established in the Delphi case, as the facts on which the judgement is based differ. He added the following reflection on the media's moral responsibility to refrain from further contaminating the public space:

> Consequently, this judgment should in no way be employed by Internet providers, in particular those who benefit financially from the dissemination of comments, whatever their contents, to shield themselves from their own liability, alternative or complementary to that of those persons who post degrading comments, for failing to take appropriate measures against these envenoming statements. If it is nevertheless used for that purpose, this judgment could become an instrument for (again!) white-washing the internet business model, aimed at a profit at *any* cost.[94]

The most recent ECHR case concerning the media's responsibility for user-generated content is Pihl v. Sweden.[95] Unlike the two previous cases, the ECHR declared the application inadmissible on the grounds that it was manifestly ill-founded. The primary reason for such a decision was that the applicant claiming to have been the victim of a defamatory online comment sued the non-profit organization on whose website the comment was published. The ECHR found that the domestic court had properly balanced the competing rights and interests in rejecting his claims. From the ECHR's reasoning, it becomes apparent that the other two facts similar to those in the Hungarian case also contributed to the finding of no violation concerning Article 10 of the Convention. First, the statement, although offensive, did not amount to hate speech. Second, it was taken down immediately upon notification from the applicant.

These cases indicate what national legislatures and domestic judicial and regulatory authorities should consider with regard to the issue of the electronic media's responsibility for user-generated content. The provision on the media's responsibility

94 Ibid.

95 ECHR Pihl v. Sweden, application no. 74742/14, 2017.

should not be used as a *carte blanche* for holding them liable for any offensive or inappropriate user-generated content (comments). This would undoubtedly result in *sui generis* censorship, contrary to the principles of free speech. On the other hand, the mere theoretical possibility of holding Internet media responsible for violations without adequate enforcement could be interpreted as a poor signal from regulators that everything is allowed for the sake of profit. Therefore, sound and fair decision making should strive to balance competing interests. In this regard, decision makers should keep the following issues in mind: the statement's content (zero tolerance for hate speech), the electronic medium's profile (small non-profit vs. large profit-oriented corporations), preventive measures taken by the media (filtering of harmful content, i.e., hate speech, incitement to terrorism, images and video-clips of the sexual abuse of children and similar content), and the promptness of the media's reaction in removing the disputed content.

8.2. Is there a need for lex specialis social network regulation?

Regarding the need to enact a special law that would regulate harmful content (including fake news) on social networks, it is important to determine whether this is necessary or whether the existing legislation is sufficient. First and foremost, there is no doubt that the legislative term 'electronic publication' (as used in the EMA) excludes social networks, which are webpages and applications that allow users to create and share content or participate in social networking. Given the previously elaborated definition of electronic publication, which includes subjects or content providers, method of implementation, and purpose, it is obvious that the concept of an electronic publication is narrower than that of the social network. The social network concept focuses on user-generated content. Unlike electronic publications, there is no emphasis on editorially designed program content published via the Internet with the purpose of informing and educating the public. Therefore, under Croatian legislation, social networks do not fall under the ambit of the legislation regulating electronic media.

Nevertheless, irrespective of the difference between the legal term 'electronic publication' and the notion of the 'social network,' which has not been legally defined (at least not in legally binding domestic legislation), it would be wrong to think that expressions and statements posted on social networks are in a legal vacuum. On the contrary, any content that conflicts with positive criminal legislation, such as hate speech, incitement to terrorism, incitement to violence, and hatred, will be prosecuted, and the perpetrator will be punished under general legislation (the Criminal Code). The same applies to the dissemination of fake news, with the difference that there will be liability for a misdemeanor and not a criminal offense. This means that in terms of criminal/misdemeanor liability, there is no difference as to whether harmful content was published in an electronic publication (e.g., an Internet news portal) or via a social network (e.g., content or commentary posted on Facebook).

As previously mentioned, some commentators have advocated German law regulating social networks as a good model for future Croatian legislation concerning social networks (see supra Section 7). As German law refers to content that has already been criminalized (a reference to the list of offenses), the new legislation is not about defining harmful content *per se*; rather, it is concerned with private companies' responsibility to filter such content and remove it from their domains. Germany was the first European country to introduce an obligation to filter harmful content on social networks. Those under the scope of the law are profit-seeking service providers that operate "Internet platforms which are designed to enable users to share any content with other users or to make such content available to the public (social networks)."[96] However, the law does not apply to all social network providers, only to those with at least 2 million registered users in Germany. They are obligated to take measures to filter, block, and remove criminalized harmful content that could be subsumed under the Criminal Code's list of offenses.[97]

There are some problems concerning the concept of social network providers' responsibility for user-generated content. First and foremost, it is a matter of transferring the responsibility for determining the issue of *prima facie* illegal content from public authorities to the private sector. According to longstanding principles and procedures in states governed by the rule of law, whether something is illegal is a matter that should be adjudicated in legal proceedings before the courts or other competent (public) authorities. The *ratio legis* for this is that any removal of content and penalization of its author must be based on law and only in cases where it is necessary in a democratic society and proportionate to the aim pursued by certain restrictions established under the law. Given that any filtering or blocking of content is a restriction of the right to freedom of expression, the weighing of protected interests must ultimately be left to the state (judiciary) and not to the private sector alone.

Furthermore, simple technical removal of inadmissible content creates a risk of impunity for the author of that content. There is a justified concern that prioritizing simple content deletion could jeopardize the justification of the criminal prosecution and punishment of those responsible for producing the content. This further raises the question of whether impunity will lead to reoffending. It is also closely related to the psychology of offenders. Imagine, for example, that it is forbidden to dispose of waste in a certain protected place (e.g., a forest). Some citizens turn a deaf ear to it and decide to dump garbage in the woods. The authorities in charge of clearing the forest remove the waste and transfer it to the appropriate disposal place, as provided by law. Had the competent authority failed to do so, they would have been sanctioned for failure. However, those who dumped the waste go unpunished. What message is being sent? Obviously, this one: Continue to dispose waste in forbidden places, and be assured that someone will clean up behind you because, otherwise,

96 Kettemann, 2019. Available at: https://bit.ly/3CxZArj.
97 Ibid.

that authority will be sanctioned. This is an instance of not properly insisting on consequences. Instead, the responsibility of those directly liable for the placement of the hypothetical waste should be strengthened through better collaboration between intermediaries and the state, not by entrusting decision making in this very delicate sphere of the most fundamental human rights to the private sector alone.

It should not be forgotten that the ECHR standards regarding protected and prohibited expression also apply to the Internet. This was clearly established in the cases of Yildirim v. Turkey[98] and Cengiz and Others v. Turkey.[99] The margin of appreciation on this is left to the state and depends on the type of expression (e.g., some poetic and satirical forms enjoy a very high threshold of protection), the mode of expression (e.g., even expressing opinions that are 'offensive' and 'shocking' will not be *a priori* prohibited if they serve a positive social function, dialogue, or pluralism in a democratic society), etc. Only certain content or forms of expression in this sense are prohibited in Europe (e.g., hate speech, Holocaust denial, incitement to discrimination, etc.).

In order to protect freedom of expression as one of the fundamental values of a democratic society, it is assumed that certain content is allowed if there are no circumstances that preclude it. In other words, the burden of proof is on the prosecutor/plaintiff or whoever claims that certain content is prohibited. The exception in this regard is defamation; the reason for the inversion of the burden of proof in this case has already been explained in Section 3. Shifting responsibility for determining whether particular content is *prima facie* illegal to the private sector alone fully relativizes freedom of expression by inversion through assuming something as *prima facie* illegal. This paradigmatic shift could be very dangerous for freedom of expression. Any reasonable service provider will, in dubious situations and faced with the risk of high penalties and reputational damage, act quite safely, preferring 'easy censorship' to the detriment of freedom of expression.

Concerning *prima facie* prohibited content, it should also be noted that this fact may be relatively easy to identify in some cases (e.g., content related to child sexual abuse). However, in other cases, it will not be that easy. For example, in cases of incitement to terrorism or public provocation to commit terrorist offenses, it will not always be *prima facie* clear whether this is a prohibited expression or one that enjoys protection under Article 10 of the Convention. The same applies to incitement to violence and hatred, and especially to fake news. Standards for distinguishing between what is allowed and what is prohibited under Article 10 of the Convention exist and have been elaborated in the ECHR's jurisprudence. However, the application of these standards in specific situations should not be left to the exclusive assessment of social network intermediaries' technical protocols and procedures. In this regard, the fact that artificial intelligence algorithms are often used to filter activities raises further legal and ethical doubts. Although these algorithms are programmed by

98 ECHR Yildirim v. Turkey, application no. 3111/10, 2012.
99 ECHR Cengiz and Others v. Turkey, application 48226/10, 2020.

humans, the operative filtering/blocking decision is taken by an algorithm fueled by artificial intelligence. This is another reason for concern about and reconsideration of the existing models (e.g., the German model).

This criticism does not mean that the private sector should be excluded from the regulation of social networks. On the contrary, a wide range of Internet intermediaries (including social network providers) must be involved in this process. However, their involvement should neither be seen nor treated as a substitute for the competent (public) authorities' balancing of competing rights and interests. The reason for this is clear. The former's role is not to protect freedom of expression, but rather to make a profit on the open market. It follows that the measures they take (filtering, content removal) lack deterrent effect in terms of special and general prevention. Hence, it is unlikely that a model relying solely on their responsibility for user-generated content would prevent the creation and dissemination of harmful content on social networks. Last but not least, law enforcement and the judiciary could misunderstand this to mean that the harm has been remedied and that no further action is needed.

That is why models of Internet intermediaries' (including social network providers') responsibility should be complementary to those involving other interlocutors, particularly those who are, per the Constitution, in charge of balancing competing rights and interests. In terms of semantics, passing new legislation on regulating Internet intermediaries' rights and responsibilities could be understood *per se* as a political tool to suppress freedom of expression on the Internet. Given that social networks constitute a global phenomenon and that the suppression of illegal and harmful content requires effective and genuine international cooperation, it would be preferable to further discuss and eventually negotiate a global (or at least regional) legal framework based on established human rights standards. In the meantime, as an alternative to unilateral legislative reform, some other measures toward better collaboration models between intermediaries themselves as well as between intermediaries and state authorities should be given preference. This could entail, for instance, adopting memoranda of understanding and codes of conduct, organizing and attending training and education for intermediaries' employees, promoting media literacy among users, etc. In any case, the standards established by the 2018 Council of Europe Recommendation on Internet intermediaries' roles and responsibilities should be closely followed to avoid interference with protective mechanisms under the Convention (by not imposing a general obligation or responsibility on intermediaries to monitor the content to which they merely provide access to, transmit, or store through the provision of an effective remedy for all human rights and fundamental freedom violations set forth in the Convention by Internet intermediaries, etc.).[100]

100 Council of Europe Recommendation on roles and responsibilities of Internet intermediaries, CM/Rec(2018)2. Available at: https://bit.ly/3zr3lgo.

9. Conclusion – Does it make sense to counter fake news in a world where the truth has (almost) disappeared?

Some commentators have suggested that countering fake news does not make any sense in a world where the truth has disappeared. They are deeply convinced that we already live in a post-truth world, that is, an environment where the common standards of humanity that were agreed upon through layers of history first faded and then vanished. For those who are more moderate, the modern world seems to be in a 'crisis of truth' or an epistemological limbo. They have suggested that ubiquitous post-modernism has relativized most historical narratives. Some modern collective ideologies (movements) have served the same purpose. The first victim of their aggressive imposition of the new narrative(s) was freedom of expression, which is on the verge of being altered for the sake of empty political correctness.

The truth is, as always, somewhere in between. While the truth has not disappeared entirely, it has indeed come under multiple attacks. The cannons are being fired not only by those already mentioned, but also by profit-oriented corporations, non-democratic governments, totalitarian regimes and ideologies, aggressive non-state actors, terrorist and anarchist cells, and many others. Cyberspace is the central arena for this global warfare. The result is a physical world polarized by emotions (instead of harmonized by reason) and divided through street spectacles (instead of united through democratic institutions). How do we get back on the right track?

The answer lies in the problem itself. The truth has to be revitalized, protected, and reinforced. It is absolutely crucial because a still dystopian (fortunately) post-truth world will lack values, as no truth means no values. A world without values will distort the concepts upon which modern (Western) societies were built, namely democracy, human rights, and the rule of law. In more practical terms, defending the truth means protecting the (generic) constitution as an expression of will and consensus on the most fundamental values shared by free citizens in democratic societies governed by the rule of law.

On this quest, the most crucial task is to keep a rational approach that prefers soft law alternatives, building partnerships, and investing in education (as a barrier to indoctrination and probably the most common method utilized by the aforementioned anti-truth/anti-democratic initiatives and alliances). With this in mind, the over-criminalization/overregulation of the Internet and social networks does not seem to be a viable model. Specifically, while it is undeniable that the negative potential of fake news is a serious threat to democratic societies, the idea that this phenomenon should be suppressed at all costs is perhaps even more dangerous. Democratic societies are based on the concept of freedom of expression, which is why a widespread campaign advocating various forms of repressive action against fake news would be deeply wrong and harmful to the very core of democracy.

Along these same lines, not all fake news threatens democracy. Rather, the threat is posed only by those massive campaigns run by the aforementioned entities

with the aim of destabilizing the system, causing panic, creating confusion, and fueling social polarization. Therefore, to avoid preventive over-filtering and over-blocking, the self-regulation of the Internet and social networks must be complementary to other measures rather than exclusive. Social media, content providers, social network providers, Internet service providers, and other interlocutors should closely collaborate among themselves as well as with regulators, law enforcement, and the judiciary. The court is the most appropriate forum (and the only one that is constitutionally authorized) to balance competing interests. Domestic court judges should follow the standards established in the ECHR's jurisprudence, which will certainly continue to evolve through new ethical and legal dilemmas concerning digital technology and artificial intelligence.

To conclude, it makes perfect sense to counter fake news as a phenomenon that harms society (created and disseminated on a large scale and/or capable of causing panic and destabilizing democratic institutions). The same applies to other forms of harmful content on the Internet and social networks. However, the truth has not disappeared from the world entirely. Sometimes it is under pressure, hiding, or silent, but those are temporary states. Like water, which is also essential for life, the truth will find a pathway.

Bibliography

ALABURIĆ, V. (1996) 'Sloboda izražavanja i kaznenopravna zaštita časti i ugleda: teorijski i praktični problemi u okolnostima tranzicije u Hrvatskoj', *Hrvatski ljetopis za kazneno pravo i praksu*, 3(2), pp. 537–557.

ALEXANDER, L. (2005) *Is There a Right of Freedom of Expression?* Cambridge University Press.

BILOŠ, A. (2020) 'Izazovi internetskog informacijskog prostora: kako stanovnici Republike Hrvatske doživljavaju problem lažnih vijesti?', *International Journal of Marketing Science*, 3(1), pp. 166–185.

CASSIM, F. (2015) 'Regulating hate speech and freedom of expression on the internet: Promoting tolerance and diversity', *South African Journal of Criminal Justice,* 28(3), pp. 303–336.

DERENČINOVIĆ, D. (2018) 'Freedom of Expression and Legal Protection of Religious Feelings in Europe — From Reconciliation to Complementarity' in Babie, P. and Savić V.I. (2018). *Law, Religion and Love – Seeking Ecumenical Justice for the Other, Routledge,* 2020, pp. 194–212.

GETOŠ KALAC, A., BEZIĆ, R. (2017), 'Criminology, crime and criminal justice in Croatia', *European Journal of Criminology,* 14(2), pp. 242–266.

ROKSANDIĆ VIDLIČKA, S. (2014) 'Possible future challenge for the ECTHR: Importance of the Act on exemption and the Sanader case for transitional justice jurisprudence and the development of transitional justice policies', *Zbornik Pravnog Fakulteta Zagrebu,* 64(5-6), pp. 1091–1120.

HENSON, B., REYNS, B. W., FISHER, B. S. (2013) 'Fear of crime online? examining the effect of risk, previous victimization, and exposure on fear of online interpersonal victimization', *Journal of Contemporary Criminal Justice,* 29(4), pp. 475–497.

KETTEMANN, M. (2019) *Follow-Up to the Comparative Study on 'Blocking, Filtering, and Take-Down of Illegal Internet Content', Germany.* Available at: https://rm.coe.int/dgi-2019-update-chapter-germany-study-on-blocking-and-filtering/168097ac51. (Accessed: 10 May 2021).

MOSLAVAC, B. (2020) *Lažna uzbuna i lažne vijesti*. Available at: https://www.iusinfo.hr/strucni-clanci/lazna-uzbuna-i-lazne-vijesti (Accessed: 10 May 2021).

MUELLER, R. (2019) *Report on the investigation into Russian interference in the 2016 presidential election (Vols. 1-2)*, Washington, DC: Academic Press.

NIELSEN, G. (2020) Populism, Fake News, and the Flight From Democracy.

DALKIR, K., KATZ, R. (2020). Navigating Fake News, Alternative Facts, and Misinformation in a Post-Truth World. IGI Global Disseminator of Knowledge.

PETERS, M. A., RIDER, S., HYVÖNEN, M., BESLEY, T. (Eds.) (2018) *Post truth, fake news: Viral modernity and higher education*, Singapore: Springer. DOI: 10.1007/978-981-10-8013-5.

ROKSANDIĆ, S. R., MAMIĆ, K. (2018) 'Zlouporaba društvenih mreža u javnom poticanju na nasilje i mržnju i širenju lažnih vijesti: potreba transplantiranja njemačkog Zakona o jačanju provedbe zakona na društvenim mrežama', *Hrvatski ljetopis za kazneno pravo i praksu,* 25(2), pp. 329–357.

SCHABAS, W. A. (2016) *The European Convention on Human Rights: A Commentary.* Oxford University Press, pp. 444–482.

SEARS, A. M. (2020) 'Algorithmic speech and freedom of expression', *Vanderbilt Journal of Transnational Law,* 53(4), pp. 1327–1376.

VOLOSEVICI, D. (2020). Brief considerations on telework during pandemic. *Jus et Civitas: Journal of Social and Legal Studies*, 7(1), 109–116.
WARDLE, DERAKHSHAN, H. (2017). Information Disorder: Toward an interdisciplinary framework for research and policy making. Strasbourg, Council of Europe.

Legal sources

Cengiz and Others v. Turkey, application 48226/10, 2020.
Constitutional Court, U-III-2858/2008, 22 December 2011.
Contribution of COMECE and CEC to Croatia's EU Council Presidency Programme "A strong Europe in a world of challenges," January 2020.
Criminal Code, Narodne novine (Official Gazette), 125/2011, 142/2012, 56/2015, 61/2015, 101/2017, 118/2018, 126/2019.
Delphi v. Estonia, application no. 64569/09, 2015.
Electronic Media Act, Narodne novine (Official Gazette) 153/2009, 84/2011, 94/2013, 136/2013.
Handyside v. the United Kingdom, application no. 5493/72, 1976.
Law on Misdemeanours against Public Order and Peace, Narodne novine (Official Gazette) 47/1990, 55/1991, 29/1994.
Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, application no. 22947/13, 2016.
Observer and Guardian v. the United Kingdom, application no. 13585/88, 1990.
Otto Preminger Institute v. Austria, application no. 13470/87, 1994.
Pihl v. Sweden, application no. 74742/14, 2017.
Prager and Oberschlick v. Austrija, application no. 15974/90, 1995.
Yildirim v. Turkey, application no. 3111/10, 2012.

Websites

https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2017&start=1990 (Accessed: 10 May 2021).
https://www.internetworldstats.com/stats.htm (Accessed: 10 May 2021).
https://www.arbona.hr/blog/drustveni-marketing/facebook-i-instagram-u-hrvatskoj-zanimljive-statistike-2019/2832 (Accessed: 10 May 2021).
https://www.weforum.org/reports/global-competitiveness-report-2019 **(Accessed:** 10 May 2021).
Constitution of the Republic of Croatia (Official Gazette) 56/1990, 135/1997, 113/2000, 28/2001, 76/2010, 5/2014; for updated version in English visit: https://www.usud.hr/sites/default/files/dokumenti/The_consolidated_text_of_the_Constitution_of_the_Republic_of_Croatia_as_of_15_January_2014.pdf (Accessed: 10 May 2021).
https://www.enciklopedija.hr/natuknica.aspx?id=11246. Accessed 10 May 2021).
Convention for the Protection of Human Rights and Fundamental Freedoms, https://www.echr.coe.int/documents/convention_eng.pdf (Accessed: 10 May 2021).

Draft Electronic Media Act, https://sabor.hr/prijedlog-zakona-o-elektronickim-medijima-prvo-citanje-pze-br-62-predlagateljica-vlada-republike?t=121735&tid=209141 (Accessed: 10 May 2021).
https://www.aem.hr/vijece/ (Accessed: 10 May 2021).
Izvješće o radu Vijeća za elektroničke medije i Agencije za elektroničke medije za 2019. godinu, https://www.sabor.hr/hr/izvjesce-o-radu-vijeca-za-elektronicke-medije-i-agencije-za-elektronicke-medije-za-2019-godinu (Accessed: 10 May 2021).
https://narod.hr/hrvatska/zasto-je-hnd-protiv-imenovanja-katje-kusec-u-vijece-za-elektronicke-medije (Accessed: 10 May 2021).
https://www.vecernji.hr/vijesti/marko-juric-emisiju-na-z1-odjavio-cuvajte-se-cetnika-s-cvjetnog-trga-1053695 (Accessed: 10 May 2021).
https://www.vecernji.hr/premium/reuters-se-ne-moze-zaustaviti-to-nisu-shvatili-ni-desni-ni-lijevi-1068920 (Accessed: 10 May 2021).
https://www.vecernji.hr/vijesti/rakic-podnijela-ostavku-na-mjesto-predsjednice-vijeca-za-elektronicke-medije-sad-mi-je-previse-1069380 (Accessed: 10 May 2021).
https://www.vecernji.hr/vijesti/nekrofilske-orgije-vrijedaju-katolike-ali-nedovoljno-za-kaznu-1085496 (Accessed: 10 May 2021).
https://www.bitno.net/vijesti/hrvatska/osnivac-index-hr-a-matija-babic-pozvao-na-paljenje-crkava-i-kidanje-glava/ (Accessed: 10 May 2021).
Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation, or administrative action in the Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities. PE/33/2018/REV/1. OJ L 303, 28.11.2018, p. 69–92. https://eur-lex.europa.eu/eli/dir/2018/1808/oj (Accessed: 10 May 2021).
https://www.hnd.hr/vanja-juric-stavlja-se-prevelika-i-pregeneralna-odgovornost-na-nakladnike (Accessed: 10 May 2021).
https://www.bug.hr/propisi/hrvatska-dobiva-novi-strozi-zakon-o-elektronickim-medijima-17557 (Accessed: 10 May 2021).
://www.collinsdictionary.com/dictionary/english/fake-news (Accessed: 10 May 2021).
Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the Committee of the regions tackling online disinformation: a european approach. com/2018/236 final. https://eur-lex.europa.eu/legal-content/en/txt/?uri=celex%3a52018dc0236 (Accessed: 10 May 2021).
https://www.britannica.com/event/Ems-telegram (Accessed: 10 May 2021).
https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation (Accessed: 10 May 2021).
https://www.poslovni.hr/hrvatska/vecina-hrvatskih-medija-bez-laznih-vijesti-o-pandemiji-4254330 (Accessed: 10 May 2021).
http://www.dorh.hr/PriopcenjaDORH-a (Accessed: 10 May 2021).
https://www.novilist.hr/novosti/hrvatska/32-prijave-za-sirenje-laznih-vijesti-o-koroni-u-hrvatskoj/ (Accessed: 10 May 2021).
https://www.vecernji.hr/vijesti/pritvoren-muskarac-koji-je-sirio-lazne-vijesti-o-potresima-1457925. (Accessed: 10 May 2021).
https://balkaninsight.com/2021/01/07/croatias-quakes-create-fertile-ground-for-fake-news/ (Accessed: 10 May 2021).
https://www.izbori.hr/site/UserDocsImages/1975 (Accessed: 10 May 2021).

https://www.jutarnji.hr/vijesti/hrvatska/mup-tvrdi-da-je-sirenje-fake-newsa-kazneno-djelo-ugledna-odvjetnica-se-ne-slaze-u-nasem-zakonodavstvu-to-je-tek-prekrsaj-10086295 (Accessed: 10 May 2021).

https://www.croatiaweek.com/croatia-to-get-museum-of-fake-news/ (Accessed: 10 May 2021).

https://mlv.hr/o-nama/ (Accessed: 10 May 2021).

https://www.medijskapismenost.hr (Accessed: 10 May 2021).

Joint communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Action Plan against Disinformation, JOIN/2018/36 final. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018JC0036 (Accessed: 10 May 2021).

Joint communication to the European Parliament and the Council, Joint Framework on countering hybrid threats a European Union response, JOIN/2016/018 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016JC0018 (Accessed: 10 May 2021).

Priorities of the Croatian Presidency of the Council of the European Union, http://euromil.org/priorities-of-the-croatian-presidency-of-the-council-of-the-eu/ (Accessed: 10 May 2021).

The effects of campaigns on participation in political decision-making, rapporteur: Marina Skrabalo, SOC/630-EESC-2019, https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/effects-campaigns-participation-political-decision-making-exploratory-opinion-request-croatian-presidency (Accessed: 10 May 2021).

https://www.reuters.com/article/eu-telecoms-int-idUSKBN2740Q3 (Accessed: 10 May 2021).

https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html (Accessed: 10 May 2021).

https://www.ombudsman.hr/hr/izrazavanje-u-javnom-prostoru/ (Accessed: 10 May 2021).

https://www.jutarnji.hr/life/tehnologija/treba-li-nam-doista-novi-zakon-o-nedopustenom-ponasanju-na-internetu-strucnjak-radi-se-o-pokusaju-ljudi-koji-nemaju-pojma-kako-funkcionira-internet-8269355 (Accessed: 10 May 2021).

https://www.gong.hr/hr/dobra-vladavina/bez-regulacije-drustvene-mreze-su-alat-za-manipula/ (Accessed: 10 May 2021).

https://hr.n1info.com/vijesti/a459328-kuca-ljudskih-prava-govor-mrznje-nije-dovoljno-ukloniti-vec-i-sankcionirati/, 10 May 2021).

Council of Europe Recommendation on roles and responsibilities of internet intermediaries, CM/Rec(2018)2, https://www.coe.int/en/web/freedom-expression/committee-of-ministers-adopted-texts/-/asset_publisher/aDXmrol0vvsU/content/recommendation-cm-rec-2018-2-of-the-committee-of-ministers-to-member-states-on-the-roles-and-responsibilities-of-internet-intermediaries?inheritRedirect=false (Accessed: 10 May 2021).

Chapter VI

Legal Aspects of Content Moderation on Social Networks in Slovenia

Kristina Čufar

1. Introduction

This chapter analyzes the existing legal framework for the regulation of content on social networks[1] and its implications for the freedom of expression and Slovenian scholarship on the subject. The goal of the chapter is to establish how public and private regulation construe the limits of the freedom of expression on social networks by focusing on the phenomena of hate speech on Facebook and the spread of misinformation (false information created and disseminated without malicious intent) and disinformation (false information deliberately created and disseminated with the intent to deceive, often referred to as fake news) on social media. Social networks are transforming the way Slovenians communicate and the way they access, create, disseminate, discuss, and perceive information. For instance, Slovenian language and grammatical rules are transformed when users express themselves online.[2] Facebook and similar platforms affect the way individuals see themselves and

1 Social networks are web-based services that permit users to open a profile or account on which they can share their personal information and opinions and establish connections and communicate with other users; social media are digital platforms for information exchange. Since platforms very often allow both communication and the exchange of information, the terms are used interchangeably in this chapter. For more, see: Boyd and Ellison 2007.

2 Fišer, Erjavec, and Ljubešić, 2016.

Kristina Čufar (2021) Legal Aspects of Content Moderation on Social Networks in Slovenia. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 175–216. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_6

relate to others.[3] The possibility to express oneself and connect with others may contribute to the empowerment of traditionally sidelined groups[4] and play an important role in the creation and coordination of social movements.[5] Social networks also pose several challenges, among them the unprecedented spread of mis- and disinformation and amplified bullying and harassment[6] owing to the new, lower standards of acceptable expression online.[7] Since independence, Slovenia has cultivated a very permissive attitude toward freedom of expression, owing to the abuse of Art. 133 of the Criminal Code of the Socialist Federal Republic of Yugoslavia prohibiting hostile propaganda.[8] Recently, however, in the light of vulgar and offensive public communication on social networks, politicians across the political spectrum are urging for stricter regulation and prosecution of hate speech.[9] Online hate speech is much discussed in Slovenia, which makes it a valuable example to study in the context.[10]

Facebook's popularity, diverse user structure, and active content moderation make it a good example on which to map the impact of social networks on content moderation and freedom of expression.[11] Among Slovenians aged 16–74, 87% use the Internet regularly,[12] 82% have at least one social network account,[13] and more than half have a Facebook account.[14] While Facebook has the most users, Instagram ranks second as the fastest growing social media platform, while Twitter is third, with 50,000 daily users.[15] Twitter is especially popular amongst Slovenian politicians and political influencers for direct communication with the public.[16] Twitter posts receive a lot of attention from traditional media (media distributing edited content, like radio, television, newspapers, etc.), indicating a reach that goes far beyond the

3 Selak and Kuhar, 2020.
4 Petrič et al., 2015.
5 Škerjanec, 2013; Prislan, 2013.
6 Završnik and Sedej, 2012; Oblak Črnič and Jontes, 2019.
7 Jereb, 2020.
8 Bajt, 2017a.
9 "Koalicija proti sovražnim napadom. Opozicija: Tudi sami morate prevzeti odgovornost." 2021.
10 ZLOvenija Tumblr page, a word play on evil ('zlo') and Slovenia ('Slovenija') exposed unprecedented increase of incendiary speech on Facebook's platform during the so-called 2015 refugee crisis. ZLOvenija published instances of hate speech against migrants expressed on public Facebook profiles and groups, along with the names and photographs of the speakers. Plesničar and Šarf 2020; Oblak Črnič 2017; A look at Slovenian Facebook ecosphere demonstrates that certain topics (like migration and LGBTQ+) attract high numbers of offensive Facebook comments. Vehovar et al., 2020.
11 Besides Facebook, Instagram, and Twitter, other popular platforms in Slovenia include TikTok (a video-sharing platform), Reddit (a discussion and content-sharing site with a lot of active Slovenian users gathering on Slovenia subreddit community), R/Slovenia, YouTube (video sharing), Snapchat (communication, content sharing), Viber (communication, content sharing), WhatsApp (communication, content sharing), Tinder (dating), Tumblr (microblogging), etc. "Družabna Omrežja | Safe.Si" 2021.
12 Statistical Office of the Republic of Slovenia, 2020.
13 Valicon, 2020.
14 Ibid.
15 Ibid.
16 Godnov and Redek, 2014.

active platform users.[17] Many traditional media organizations also operate social media profiles. Social media companies attract advertisers by regulating, removing, or promoting the content appearing on their digital platforms through manipulation of their users' emotional responses;[18] traditional media, whose existence depends on advertising, are struggling to compete.[19]

The chapter situates existing Slovenian regulation in the European legal framework and considers the regulative approaches of social media companies on the example of Facebook in order to demonstrate the complexity of content moderation. The chapter firstly presents Facebook's rules and procedures for content moderation to exemplify the private regulation of expression on social networks. Then, the phenomenon of so-called fake news and its perception is unpacked upon the examples of different types of mis- and disinformation in the Slovenian (social) media sphere in order to break the phenomenon down into a classification of different types of dubious or manipulated information that commonly appear. Legislation regulating the activities of traditional media is briefly considered to illustrate the different legal regimes governing traditional and social media. Legal liability for creation and dissemination of mis- and disinformation is not systematically regulated in Slovenia, yet such activities may result in civil or criminal liability. While users cannot legally demand the reinstatement of a post, they may demand the removal of an illegal post. The chapter reviews the relevant Slovenian constitutional, administrative, criminal, and civil legal norms as well as case law involving problematic user-generated content. The chapter concludes with a brief discussion about open challenges and a consideration of regulative attempts in other countries.

Whether we call limiting the freedom of expression censorship or content moderation, whether it is performed by a corporation or a state, it provokes discomfort. This chapter demonstrates that content moderation is a double-edged sword – both necessary and pernicious. The root of this paradox lies beyond the scope of legal regulation, beyond the issues of 'private' and 'public'—in the very complexity of the postmodern world. What ought to be the limits of free expression, who and according to what procedures ought to decide on these limits, etc., are political and ideological questions. The issues discussed by the chapter are by no means unique or limited to Slovenia. Most Slovenian academic literature on content moderation online adopts a global and/or European Union (EU) law perspective, indicating that Slovenia mostly follows transnational regulatory trends[20] and stressing that the regulatory challenges of the digital age ought to be addressed on a transnational level.[21] There are no easy answers when it comes to the regulation of expression and news on social networks, yet regulation is necessary, if always imperfect.

17 Mance, 2014.
18 Bakir and McStay, 2018.
19 Bašić Hrvatin, 2020.
20 E.g., Damjan, 2017; Damjan, 2019; Weingerl, 2020; Selinšek, 2015.
21 Selinšek, 2015.

2. Content moderation of inappropriate speech on social networks

Most social networks make profits through commercial use of users' data. Most social networks would prefer not to moderate content but are forced to for a variety of reasons: protection of users, removal of illegal content, and appeasing the public and existing and potential users, partners, and advertisers.[22] Social networks' market-oriented practices and the lack of democratic oversight often lead to questionable decisions.[23] Nevertheless, the question is not how to stop social networks from moderating content, but how to regulate this moderation through the entities deciding on content removal, the checks and balances in place, the means of granting democratic participation and oversight, etc. The term 'private censorship' is often used to criticize social networks' controversial decisions to take down user-generated content. However, some scholars hold that the term 'censorship' implies the state-guaranteed right to speak that social networks are not legally bound to grant and by virtue of which 'censored' users are not completely silenced in online debates, as they are free to join other social networks or create new accounts.[24] Censorship in the wide sense may be understood as any official control over the flow of ideas,[25] but this term is morally loaded and has a negative connotation; it might invoke the impression that all user-generated content ought to be permitted online. This is hardly the case: removals of child pornography, serious harassment and threats, depictions of extreme violence and cruelty, terrorist propaganda, etc., are rarely (if at all) described as an infringement of freedom of expression or censorship. The term 'content moderation' (a set of governance mechanisms intended to structure participation in debates, facilitate cooperation, and prevent abuse)[26] seems more appropriate and nuanced, as it draws attention to the complexity of the issue, placing it within the wider phenomenon of postmodern global governance. This section zooms in on the problem of hate speech online and Facebook's regulative framework. In the Slovenian context, hate speech appears in two types: illegal hate speech constituting a criminal offense, which is hate speech in the narrow sense, and hate speech in the wider sociological sense.[27] Hate speech in the wider sense refers to all instances of discriminatory speech based on the idea that certain groups of human beings are inferior to others; it is not necessarily illegal, but it is widely considered

22 Gillespie, 2018, pp. 6–24.
23 Tushnet, 2019.
24 Gillespie, 2018, p. 176.
25 Režek, 2010.
26 Grimmelmann, 2015.
27 E.g., Splichal, 2017; Završnik and Zrimšek, 2018; Jalušič, 2019; Lindič, 2017; Zobec, 2019; Teršek, 2018.

as vulgar and inappropriate.[28] Facebook's definition of hate speech corresponds to hate speech in the wider sense.[29]

2.1. Freedom of expression on Facebook in Slovenia

Media organizations actively moderate content on their sites as a part of their editorial policy and are liable for this content. The Mass Media Act (Zakon o medijih – ZMed) defines media as newspapers, magazines, radio and TV programs, electronic publications, teletext and other forms of daily or periodical publishing of *edited content* via text, voice, sound, or image available to the public (Art. 2). The definition does not include social networks that only provide platforms for user-generated content without creating or editing content themselves. As a private non-media company, Facebook is free to moderate user-generated content as it sees fit.

Slovenia does not have an official state censorship body, but certain categories of expression are prohibited under Slovenian law, as will be reviewed later. When it comes to hate speech, both the state (e.g., Human Rights Ombudsman, Advocate of the Principle of Equality, and diverse governmental campaigns[30]) and civil society organizations[31] are raising awareness. For instance, Spletno oko, which is active within the Safer Internet program (Department for Research at Centre for Social Informatics at the Faculty of Social Sciences, University of Ljubljana), allows users to report hate speech and sexual abuse of children online.[32] It has the status of a 'trusted flagger,' conferred by social networks on trustworthy organizations and individuals who frequently and accurately flag problematic content.[33] Spletno oko evaluates whether reported content might be illegal and may report it to the authorities or social networks without moderating the content.[34]

2.2. Facebook's regulatory framework

Facebook is a powerful global actor often compared to a state.[35] It is a private company that concentrates power and decision making by uniting law making, executive and quasi-judiciary power, and the power of the press.[36] It is the largest social network in the world[37] and it also owns the WhatsApp messaging service and the social network Instagram.[38] Facebook is working hard to present itself as a socially

28 Bajt, 2017b.
29 Facebook, 2021a.
30 E.g.:"Kampanja Ne sovražnemu govoru | GOV.SI" 2021.
31 E.g.: "Z (od)govorom na sovražni govor – ZaGovor" 2021.
32 "Trditve in Dejstva o Spletnem Očesu | Spletno Oko" 2021.
33 Ibid.
34 "Sovražni govor na spletnih družbenih omrežjih v Sloveniji" 2021.
35 Chander 2012.
36 Kadri and Klonick, 2019.
37 Facebook has at least 2.7 billion users. "Most Used Social Media 2020," 2021.
38 "The Facebook Company Products | Facebook Help Center," 2021.

responsible enterprise capable of balancing the fine line between the freedom of expression and guaranteeing a safe space to its users.[39] It has recently adopted the Corporate Human Rights Policy and committed itself to regular reporting to prove its commitment to human rights.[40] Content moderation is central to this process. To create a Facebook account, a user must agree to the Terms of Service[41] and thereby accept the Community Standards,[42] which are described as "a comprehensive set of policies that help […] create the conditions so people feel comfortable expressing themselves by balancing the values of voice, authenticity, safety, privacy and dignity."[43]

2.2.1. Community Standards

Facebook was founded in 2004 to target university students, but its user base quickly grew and diversified.[44] Until 2008, Facebook had no content moderation policy, only a few dozen people guided by a single page document and their instincts.[45] Facebook's growth demanded standard setting for its diverse global 'community,' resulting in globally applicable guidelines reflecting a narrowed version of the US conception of the freedom of speech;[46] EU law, individual European states' national legislation, and public pressure fueled by a variety of scandals were also important influences.[47] The Community Standards were developed and published in 2008, but Facebook's internal rules governing content moderation only became public in 2018.[48] The motivation for Facebook's content moderation is profit-oriented – the more time people spend on Facebook, the more ads are displayed to them and the more money is made.[49] It is thus in Facebook's interest to ensure that its users feel comfortable and safe while enjoying its services.

Community Standards divide problematic content into five parts: violence and incitement (coordinating harm, publicizing crime, credible threats, etc.); safety (child sexual exploitation, abuse, and nudity, glorification of suicide and self-injury, etc.); objectionable content (hate speech, adult nudity and sexual activity, etc.); integrity and authenticity (fake accounts, spam, etc.); and respecting intellectual property (copyright and trademark violations, etc.).[50] The Community Standards offer some insights into the interpretation of its provisions. For instance,

39 B. J. Johnson, 2016.
40 Facebook, 2021.
41 "Facebook: Terms of Service," 2021.
42 "Community Standards | Facebook," 2021.
43 "Community Standards Enforcement," 2021.
44 Brügger, 2015.
45 Klonick, 2020.
46 Klonick, 2017.
47 Ibid.
48 Bricket, 2018.
49 Klonick, 2017.
50 "Community Standards | Facebook," 2021.

photographs of female nipples are generally not allowed but may appear in the context of breastfeeding or post-mastectomy awareness-raising; sculptures and other artistic depiction of nude figures are also allowed; glorification of suicide and self-injury is not allowed but sharing experiences and raising awareness about these issues is permitted; etc.[51]

Facebook detects potential violations through reports from trusted flaggers, ordinary users, and artificial intelligence (AI).[52] Flagged content is evaluated according to the order of priority decided by the AI. Removal decisions are sometimes fully automated. According to Facebook, a large percentage of inadmissible content is removed by AI before users see it.[53] Facebook may sanction the breach of rules by removing the post, disabling the account, covering content with a warning, and reporting all apparent instances of child exploitation to the National Center for Missing and Exploited Children. If illegal activity is suspected, Facebook alerts the police. Facebook admits that the process is not entirely smooth: "In some cases, we make mistakes because our policies are not sufficiently clear to our content reviewers […] we make mistakes because our processes involve people, and people are fallible."[54] If a user does not agree with Facebook's decision, they may request a review. Facebook takes another look at the case, usually within 24 hours.[55] If the review finds that Facebook made a mistake, the user is notified and their post restored or access to the suspended account enabled.[56]

The mistakes that occasionally occur in the content moderation process are best illustrated by the scandal caused by Facebook's removal of the iconic 'Napalm Girl' photograph.[57] The image depicting a naked Vietnamese girl escaping a napalm attack during the Vietnam War breaks the rules about child nudity although it is not pornographic and is rather a famous historical image. This case is by no means Facebook's only controversial content moderation decision, and it reveals just how complex the interpretation and enforcement of Community Standards can be. Facebook's content moderation is rightfully criticized for lacking transparency, oversight, and democratic participation.[58] Considering Facebook's power, several issues repeatedly arise: the freedom of expression (transparency, due process, democratic oversight, etc.); the safety, privacy, and dignity of users targeted by other users' speech; national and transnational legislation with which Facebook is bound to comply; Facebook and its users' criminal and civil liability; Facebook's questionable content moderation decisions; etc.

51 Ibid.
52 King and Gotimer, 2020.
53 Ibid.
54 Bricket, 2018.
55 "I Don't Think Facebook Should Have Taken down My Post. | Facebook Help Center," 2021.
56 "My Personal Account Was Disabled | Facebook," 2021.
57 Ibrahim, 2017.
58 Heins, 2013.

2.2.2. Oversight Board

Public pressure to make Facebook's content moderation and the underpinning rules more transparent and democratic resulted in the creation of a global body of experts independent from Facebook, namely the Oversight Board, in 2020.[59] When the Board's trust, charter, and bylaws were being prepared, Facebook's founder and CEO Mark Zuckerberg described the body as an equivalent of the Supreme Court.[60] Users can appeal Facebook's content moderation decisions to the Board, and Facebook is bound by its decisions. Before appealing to the Board, the user must exhaust Facebook's internal appeals.[61] The Board is a new body and it is difficult to assess how it will influence the industry, nation states, and freedom of expression.

The Oversight Board's bylaws[62] are similar to traditional corporate and nonprofit bylaws and define the arrangement between the Board, Facebook, and the Oversight Board Trust, as well as the role of Facebook users.[63] The Board is composed of experts and civic leaders from around the globe, and it has discretion over the cases it chooses to hear – it is supposed to review the toughest cases with significant real-world impact.[64] The Board may also hear the cases of users who reported problematic content that was not removed. Facebook may also refer cases; for an (in)famous example, the indefinite suspension of former US president Trump's Facebook and Instagram accounts that followed the January 6 Capitol invasion was referred to the Board.[65] The Board upheld Facebook's decision, but also criticized the indeterminate penalty, demanding that Facebook review it.[66] The Board also recommended several actions Facebook should take in order to ensure more transparent procedures.

2.2.3. Potential problems for Slovenian Facebook users

Facebook's Community Standards are not translated in Slovenian.[67] Since some of Facebook's Slovenian users do not speak English, the omission of translation alone raises questions about transparency. Despite Facebook's reassurance about its technology's great efficiency and sophistication, concerns that AI may be arbitrary and lack certain traits and nuances of human reasoning might also be problematized. To illustrate one set of problems that might arise from ignoring Facebook users'

59 B. Harris 2020b.
60 Klonick 2020.
61 "Oversight Board | Independent Judgment. Transparency. Legitimacy." 2021.
62 "Bylaws – Oversight Board" 2021.
63 B. Harris 2020a.
64 Facebook's involvement in choosing the original Board members (who are supposed to independently choose future members) is one of the many potential flaws in the process of creating the Oversight Board. Klonick 2020
65 "Referring Former President Trump's Suspension From Facebook to the Oversight Board" 2021.
66 Oversight Board 2021.
67 Facebook 2021b.

linguistic diversity, the 2018 genocide in Myanmar serves as a chilling example. The incitement of violence against the Rohingya ethnic minority on Facebook played a considerable part in the tragedy.[68] Following the tragedy, Facebook's role in the genocide was scrutinized, demonstrating that Facebook was the primary source of news for 40% of Myanmar's population and only four content reviewers spoke Burmese at the time.[69] Today, Facebook employs human reviewers fluent in over 50 languages[70] that supplement the AI and bring the human touch and understanding of contexts and cultural norms.[71] The fact that Facebook's rules are not translated may hold consequences for users who only speak Slovenian. Not only are they not able to familiarize themselves with the Community Standards, their ability to challenge Facebook's removal of their posts is severely limited, especially considering that even English-speaking users describe Facebook's appeal process as 'speaking into the void.'[72]

While users might not be included in the creation and implementation of Community Standards, the pressure media and civil society exert does influence Facebook's platform governance. Facebook is not as unbound in its sovereignty as it might seem and is entering into complex relationships with states and their organizations.[73] The regulation of expression on social networks is a complex power struggle between states and multilateral corporations.[74] States are setting and enforcing the rules governing the freedom of expression in collaboration and through confrontation with private companies like Facebook. Traditionally, the regulation of speech and expression rested in the hands of the states directly regulating publishers and speakers, which may be described as the direct speech regulation. This is to be distinguished from the indirect speech regulation, which targets digital infrastructure through indirect regulation.[75] The indirect speech regulation complements the direct regulation's traditional toolbox and it entails cooperation or cooptation between the public state power and private companies, collateral censorship where states target

68 Galvan 2020.

69 Yue, 2019; Some researchers nevertheless suggest that Facebook's undisputed role in the ethnic cleansing in Myanmar might have been somewhat exaggerated in Western media, see e.g.: Whitten-Woodring et al., 2020; Following the public outcry and United Nations investigation, Facebook employed over 100 reviewers fluent in Burmese. Su, 2018.

70 Supposedly, Slovenian is one of these languages, but Facebook's policy is not to reveal their number or any details pertaining to content moderation in a specific country/language.

71 Silver, 2018.

72 Vaccaro, Sandvig, and Karahalios, 2020.

73 For example, Facebook changed its Terms and Conditions in 2019 in order to make its usage of users' personal data more clear, following negotiations with the European Commission. European Commission, 2021a

74 The trade association Computer & Communication Industry Association (CCIA Europe) representing Facebook fiercely criticized the EU's proposal that Internet platforms should use upload filters as an imposition of broad private censorship. Greenfield, 2018; Nevertheless, Facebook has been using upload filters since 2015. Masnick, 2015.

75 Balkin, 2014.

users/speakers through infrastructure providers, and the private governance of companies that govern their users' online behavior.[76]

3. Mis- and disinformation on social media in Slovenia

Fake news (fabricated information in the form of news with the intention to deceive the audience) is widely used and abused, open-ended, and politically-loaded expression. 'Fake news' exploded into a global buzzword following Brexit and the US elections in 2016.[77] Nevertheless, the phenomenon is far from new. The invention of the press and the spread of literacy accelerated the spread of (fake) news,[78] while the steam engine gave it another boost: the Great Moon Hoax recounting life on the moon based on fake interviews, pictures, and misleading headlines published in *The Sun* in 1835 is often referred to as a quintessential example of modern fake news.[79] According to Eurostat, Slovenians are concerned about fake news: 75% of the respondents encounter mis- and disinformation at least several times a month if not daily, only 29% of Slovenian respondents trust news on social media compared to the 75% who trust the radio, 89% estimate that mis- and disinformation constitute a problem in Slovenia, and 86% believe that it is a problem for democracy in general.[80] Instances of mis- and disinformation may be broken down into several categories. This chapter provisionally organizes the problem of dis- and misinformation on social media digital platforms in Slovenia into seven categories: fake news, misinformation, conspiracy theories, satire, clickbait, political astroturfing, and deepfakes.

3.1. Classification of mis- and disinformation

3.1.1. Fake news

Fake news has been discussed as a serious issue in Slovenia for a long time.[81] The definition of fake news is open-ended and constructed on the basis of foreign literature: Fake news generally denotes fabricated news stories created with the aim of deception. The terms 'disinformation' and 'manipulation' are also used in the context. The trend of dismissing any unfavorable news as fake news, initiated by

76 Balkin, 2018.

77 'Fake news' was selected as the Collins Dictionary's official Word of the Year for 2017; while 'post-truth' was the Oxford Dictionaries Word of the Year 2016. Hunt, 2017; "Oxford Word of the Year 2016 | Oxford Languages," 2021.

78 Burkhardt, 2017.

79 Bossaller et al., 2019.

80 "Flash Eurobarometer 464: Fake News and Disinformation Online," 2018.

81 E.g.: Jančič 2017; Jontes 2010; Jukovič 2017; M. Milosavljević 2016; Vidmajer 2017.

former US president Trump, has been noticed in Slovenia too; in such cases, the term 'fake news' is (ab)used to discredit media reporting without presenting arguments or evidence that would counter it.[82] The problem of fake news in Slovenia may be illustrated in an example: In March 2021, two Twitter accounts posted an altered image of the survey results presented on a commercial TV station, portraying higher levels of support for the political parties of the ruling coalition than the original survey published on television.[83] Twitter posts included the logos of the media house and the company that conducted the survey, claiming that the results shown on TV were falsified. Many Twitter users, including prominent politicians, re-tweeted the post in the following hours. Both the media house and the market research company denied the claims as completely ungrounded and have pressed criminal charges against an unknown perpetrator. Anonymity online makes it difficult to prosecute fake news, and at least one of the accounts that originally posted the modified survey has been proven to be fake and involved in political astroturfing on a regular basis.[84] The example demonstrates the blurriness of the proposed categories of disinformation, as it includes elements of fake news, astroturfing, and conspiracy theory.

3.1.2. Misinformation

Unlike fake news, misinformation is not created and disseminated with the purpose to deceive, it is a product of the negligent spread of information that was not fact-checked. For example, in December 2020, Slovenia was eagerly anticipating the approval of the first novel coronavirus (Covid-19) vaccine, a potential beginning of the end of the pandemic. On the other hand, public distrust toward vaccination has been growing in recent years. A Facebook post by a Slovenian gynecologist claiming that the Covid-19 vaccine causes infertility spread like wildfire on social media. The story was soon debunked by experts calling attention to the lack of scientific evidence, labeling the story a conspiracy theory and fake news.[85] The gynecologist who posted the claim soon apologized, explaining that he misunderstood the title of an online article in English.[86] Nevertheless, the seed of doubt was planted, adding to the existing concerns and doubts about the rapidly developed vaccines.

3.1.3. Conspiracy theories

Like fake news, conspiracy theories have a long history and have been amplified by the rise of social media.[87] These theories are based on the idea that shadowy elites

82 Ross and Rivers, 2018; "Janša: Lažne novice se v Sloveniji širijo tudi v osrednjih medijih," 2020.
83 Pušnik, 2021.
84 Voh Boštic, 2021.
85 See some reports in traditional media: Pavlin, 2021; "Strokovnjaki: Cepivo proti covidu-19 ne povzroča poškodb posteljice," 2020; "Mit o cepivu razkrit, ni nevarnosti za neplodnost," 2020.
86 Šašek, 2020.
87 A. Zupančič, 2020.

are malevolently manipulating reality from behind the scenes.[88] Conspiracy theories thrive in troubled times, as they offer simplistic explanations for pressing problems.[89] Some global conspiracy theories have adherents in Slovenia. Supporters of anti-vaccination conspiracy theories, which have been recognized as especially dangerous to public health,[90] often congregate on social media.[91] In real life, these pressure groups are organizing protests against obligatory vaccination.[92] Several social media groups and influencers are spreading disinformation about Covid-19 and the related protective measures.[93] Their efforts are notable in real life, as these groups organize protests and other activities.[94] The global QAnon conspiracy theory, perceived as a motivating force behind the January 6 Capitol invasion, also has its adherents in Slovenia. Even the Slovenian prime minister has re-tweeted QAnon content.[95] The QAnon movement propagates, amongst other things, a belief that a satanic cult of politicians and celebrities deals with human sacrifice and pedophilia.[96] Anti-Semitic conspiracy theories like QAnon, inspired by the ancient conspiracy theory about Jews drinking the blood of Christian babies, have famously shaped the course of global and Slovenian history: Despite the small Jewish population in Slovenian territory, anti-Semitic conspiracy theories have been continuously and systematically abused for political mobilization since at least the 19th century.[97] Even a former Slovenian European Court of Human Rights (ECtHR) judge habitually spreads such conspiracy theories about the billionaire George Soros.[98]

3.1.4. Satire

Satire involves irony, exaggeration, and humor in order to expose the absurdity or stupidity of a situation or a statement. Satire is not meant to be misleading; rather, it is a form of political commentary and critique. As such, satire is not seen as the distribution of mis- or disinformation but as essential to democratic society: Satire's special status is well established in criminal (e.g., II Kp 49761/2015) and civil case law (e.g., I Cp 1206/2015). Nevertheless, when taken uncritically, satire may bleed into misinformation. In 2010, Slovenian media uncritically translated and

88 European Commission, 2021b.
89 Abram and Grušovnik, 2021.
90 Germani and Biller-Andorno, 2021.
91 E.g., Facebook groups: "Združenje Za Naravni Razvoj Otrok | Facebook" 2021; "Skupaj Za Zdravje Človeka in Narave | Facebook," 2021.
92 "Protest proti zakonu, ki bi prepovedal vpis necepljenih otrok v javne vrtce" 2018; "Nasprotniki protestirali proti obveznemu cepljenju. Strokovnjaki: ker določenih bolezni ne vidimo, se ne zavedamo nevarnosti," Facebook groups.
93 E.g., "Maske Dol | Facebook," 2021.
94 "V Mariboru večstoglava množica vzklikala: 'Maske dol, vlada pa v zapor,'" 2021.
95 Savič, 2020.
96 Hannah, 2021.
97 Pelikan, 2015.
98 E.g., Zupančič, 2020.

distributed the news story about a drunken Serb who fell on and thereby killed a shark in an Egyptian resort.[99] As it turned out, the original source of the news was a Serbian satirical webpage called Njuz.[100] Ironically, Nujz was created as a response to the spread of fake news and is publishing made-up stories to entertain and provoke critical thinking.[101] This misinformed news story eventually spread across foreign media beyond the Balkan region.[102]

3.1.5. Clickbait

Clickbait refers to flashy and exciting titles constructed to attract attention and generate clicks.[103] Usually, an attractive title is followed by an ordinary news story that does not necessarily contain mis- or disinformation. Since people often only read the titles and tend to receive information rather uncritically, clickbait may nevertheless contribute to the general spread of mis- or disinformation and may breach personal rights.

3.1.6. Political astroturfing

Astroturfing, falsely presenting ideas as originating in a grassroots movement when they are in fact launched by an organization, originates in marketing and is increasingly present in politics.[104] Astroturfing may be described as creating a false public and manipulating public opinion.[105] While the phenomenon itself is not new, the rise of social media facilitated the creation of fake accounts and the spread of disinformation on a new scale.[106] Fake accounts are sometimes automatized (bots) and sometimes operated by humans. Twitter stands out as a social network with a large percentage of fake accounts mobilized for political purposes, despite the active removal of such accounts.[107] It is estimated that the majority of Slovenian political parties utilizes astroturfing – the problem was revealed when a Slovenian MP mistakenly continued a fake-profile tweet from her official account.[108] Journalists investigated the phenomenon of fake profiles on the Slovenian Twitter scene and found that a substantive part of 'public' opinion on Twitter is generated by fake accounts using stolen or automatically generated photographs and false identities.[109] Their tweets are often re-tweeted by politicians and even presented as sources for dubious

99 "Srb v Šarm el Šejku ubil smrtonosnega morskega psa," 2010.
100 N. Milosavljević, 2010.
101 Jovanović, 2020.
102 Bates, 2010.
103 Pohar, 2021.
104 Kovic et al., 2018.
105 "Astroturfing," 2021.
106 Russian interference in the US elections exposed the problem, see generally: Karpan, 2018.
107 Hutchinson, 2020.
108 Lotrič, 2021.
109 Voh Boštic, 2021.

news stories distributed by politically affiliated media. Such media stories are then tweeted and re-tweeted, creating a circular circuit of disinformation.[110]

3.1.7. Deepfakes

Deepfakes or synthetic media are highly convincing audio files and/or videos fabricated by the AI technology generative adversarial networks (GAN).[111] New videos depicting events that never took place are generated based on actual images and videos. Implications vary from the most intimate (e.g., computer generated involuntary pornography and the related exploitation and intimidation)[112] to geopolitical (e.g., manipulations of voters or incitements of angry mobs).[113] Technology to detect deepfakes is available,[114] but deepfakes are rapidly evolving and adapting. Deepfakes are likely to strongly influence politics, journalism, and news production in the years to come. While some are preoccupied with the potential of deepfake news for political manipulations, others stress the even more concerning effects of deepfakes on people's perceptions of reality.[115] Awareness that nothing, not even video footage, can be trusted, might further contribute to the decrease of trust in the news. The growing uncertainty might contribute to general indeterminacy and cynicism.[116] A notable deepfake story is yet to break in Slovenia – for now, engagement with deepfakes is limited to stories from abroad, usually involving foreign politicians and celebrities.[117]

4. Legal regulation of communication and information on digital platforms

Scholars estimate that both Slovenian and EU media regulation constantly remain a step behind social network corporations.[118] Cooperation between state and non-state actors is favored over state intervention, indicating a privatization of regulation.[119] As mentioned, social platforms are not media organizations under the Slovenian Mass Media Act. The Act imposes a number of obligations on traditional

110 This practice is not entirely new, but it has been increasing in the recent years. Mance, 2014.
111 Chesney and Citron, 2018.
112 D. Harris, 2018.
113 Chesney and Citron, 2019.
114 E.g., Li, Chang, and Lyu, 2018.
115 Vaccari and Chadwick, 2020.
116 Završnik, 2018a; Završnik, 2018b.
117 e.g., Cijan, 2020; Gorenšek, 2019; Jenko, 2018.
118 Smokvina and Pavleska, 2019.
119 Ibid.

media and prescribes monetary fines for violations thereof (Art. 129–148b). The Mass Media Act is based upon the principles of the protection of the Slovenian language (Art. 5), freedom of expression (Art. 6), freedom to disseminate foreign media content (Art. 7), and the prohibition of the encouragement of inequality and discrimination (Art. 8). Distributers of media content and their editors must be registered in Slovenia (Art. 10). The Mass Media Act obliges media organizations to publish emergency messages (Art. 25), limits advertisement and prohibits certain forms of advertising (Art. 46-51), mandates protection of children and minors against pornography and violence (Art. 84), and stipulates the right to correction and response (Art. 26-44), which will be explored later on. Media ownership is regulated with the aim of achieving pluralism and diversity (Art. 56-63). Traditional media is rather rigidly regulated, while Internet service providers like digital platforms play by the rules set forth in the Electronic Commerce Market Act (Zakon o elektronskem poslovanju na trgu – ZEPT) that transposed Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the internal market (e-Commerce Directive) into Slovenian legal order. This section reviews the relevant legislation and case law involving potentially illegal user-generated content.

4.1. Human rights and fundamental freedoms

4.1.1. Transnational law

Freedom of expression is enshrined in all the relevant regional human rights documents, including Art. 10 of the European Convention on Human Rights (ECHR) and Art. 11 of the Charter of Fundamental Rights of the European Union. The ECtHR finds freedom of expression crucial to democratic society. Nevertheless, freedom of expression may be subject to formalities, conditions, restrictions, or penalties (Art. 10 ECHR) that should be construed strictly and be convincingly explained (following the three tests: the lawfulness of the interference, its legitimacy, and its necessity in a democratic society).[120] Limitations of the freedom of expression are likewise set by these documents and are typically found in the right to respect for private and family life (Art. 8 ECHR; Art. 7 of the Charter), protection of personal data (Art. 8 of the Charter), and the prohibition of abuse of rights (Art. 17 ECHR). The ECHR and the Charter oblige public authorities to guarantee the freedom of expression, while no such obligation can be imposed on private companies like Facebook.

When it comes to privacy and personal data protection, the EU is a trailblazer: The General Data Protection Regulation (GDPR) is a set of the toughest data privacy laws in the world. The GDPR imposes obligations on organizations anywhere if they process the personal data of EU citizens or residents, threatening high fines in the case of noncompliance. The GDPR aims to create consistent protection of personal data across the EU member states and uniform data security law. Slovenia is the only

120 Council of Europe, 2020.

EU country yet to implement the GDPR.[121] The existing Personal Data Protection Act (Zakon o varstvu osebnih podatkov – ZVOP-1) is supposed to have been amended, but the publicly available draft of the new Personal Data Protection Act (Zakon o varstvu osebnih podatkov – ZVOP-2) is yet to be discussed by Slovenian parliament. Nevertheless, as an EU regulation, the GDPR is binding and directly applicable and does not require any action on the part of Slovenia.

The EU is also dedicated to the eradication of illegal hate speech. The Council Framework Decision 2008/913/JHA on combating certain forms and expressions of racism and xenophobia by means of criminal law binds member states to ensure that public inciting to violence or hatred against certain groups and public condoning, denying, or grossly trivializing crimes of genocide, crimes against humanity, and war crimes are punishable offenses according to the member state's criminal law. Since social networks are not bound by the human rights instruments, the European Commission, Facebook, Microsoft, Twitter, and YouTube agreed to the Code of Conduct on Countering Illegal Hate Speech Online in 2016, in the wake of the 2015 terrorist attacks in France. Instagram, Snapchat, Dailymotion, Jeuxvideo.com, and TikTok have joined the Code since.[122] In the Code, the companies pledge their responsibility to promote and facilitate freedom of expression worldwide and commit to tackling illegal hate speech online by setting up processes to review notifications regarding illegal hate speech on their platforms, encouraging the flagging of problematic content, promptly responding to removal notifications, training their staff, and sharing best practices. A network of organizations conducts the regular monitoring of the Code's implementation across the EU. According to the last monitoring, the companies assess 90% of flagged content within 24 hours and 71% of the content deemed illegal hate speech is removed as a result.[123] While the Commission considers the Code "a success story when it comes to countering illegal hate speech online,"[124] it remains controversial. Several important nongovernmental organizations (NGOs) and scholars have severely criticized it for reinforcing tech companies' power to decide on the (il)legality of expression, which might lead to excessive content removal.[125]

The situation is similar when it comes to the identification and spread of mis- and disinformation – technological giants have pushed hard for a self-regulation model in the past.[126] The Code of Practice on Disinformation – agreed upon by the platforms, leading social networks, advertisers, and the advertising industry – is an example of such practice.[127] Facebook, Twitter, Mozilla, Google, Microsoft, and TikTok have joined the Code.[128] Thus, the industry has voluntarily agreed to a set of worldwide

121 Horvat, 2020.
122 European Commission, 2021c.
123 Reynders, 2020.
124 Věra Jourová in European Commission, 2021c.
125 Kaye, 2019.
126 Sánchez Nicolás, 2020.
127 European Commission, 2021a.
128 Ibid.

self-regulatory standards to fight disinformation and committed to periodic monitoring. The European Commission plans to substitute the Code with the European Democracy Action Plan based on three pillars: promoting free and fair elections, strengthening media freedom and pluralism, and countering disinformation.[129] The Action Plan is supposed to be implemented by the next European Parliament elections in 2023.

In addition to hate speech and mis- and disinformation, the EU aims to remove other types of problematic online content. The EU Directive 2017/541 on combating terrorism demands that terrorism-related online content be removed or blocked. The EU Directive 2011/93/EU on combating the sexual abuse and sexual exploitation of children and child pornography demands that such materials be removed or blocked. The EU Directive 2019/790 on copyright and related rights is introducing new obligations for Internet service providers regarding user-generated content that violates copyright and is criticized as a dangerous incentive for private censorship, indirectly pushing providers to actively monitor user-generated content.[130]

4.1.2. Slovenia

The freedom of expression is enshrined in Art. 39 of the Slovenian Constitution (Ustava Republike Slovenije – URS). It guarantees the freedom of expression of thought, speech, and public appearance, of the press, and other forms of public communication and expression. Constitutional limits of the freedom of expression are to be found in the constitutional rights of others, like the right to personal dignity and safety (Art. 34) or the right to privacy and personality rights (Art. 35). Prohibition of incitement to discrimination and intolerance and prohibition of incitement to violence and war (Art. 63) forbid any incitement to national, racial, religious or other discrimination; the inflaming of national, racial, religious or other hatred and intolerance; or any incitement to violence and war as unconstitutional, establishing the bases for the definition of illegal hate speech in criminal law. When freedom of expression clashes with the rights of others, the Slovenian Constitutional Court looks up to the ECtHR and employs the balancing of rights (e.g., decisions Up-614/15 and Up-407/14).[131]

When deciding cases involving alleged mis- and disinformation, courts must establish the appropriate balance between the freedom of expression, which includes freedom of the press and public communication, and other rights. The Slovenian Constitutional Court generally favors and protects the freedom of press. Even exaggerated and offensive statements have their place in democratic debate and serve public interest – journalists may only be found liable if they know that their reporting is based on a lie or in cases of gross negligence (Up-1019/12). The Court

129 Ibid.
130 Damjan, 2019.
131 See also: Teršek, 2019.

follows the criteria for restricting the freedom of expression of media developed by the ECtHR taking into account the contribution to public debate; whether the injured party is a public personality; prior actions of the injured party; the method of gathering of information, its correctness and context; the manner and consequences of publication; the gravity of sanction; and the differentiation between value judgments and facts (e.g., Up-1019/12; Up-417/16). Despite the high level of freedom of press granted by the Constitutional Court, overtly sensational clickbait titles that distort the facts may be considered independently of the news story they head (Up-530/14).

4.2. Service providers' liability for user-generated content

4.2.1. European Union legislative framework

Internet intermediaries' ("a wide, diverse and rapidly evolving range of service providers that facilitate interactions on the internet between natural and legal persons")[132] civil and criminal liability for user-generated content fall under the basic legal framework for information society services in the EU – the e-Commerce Directive. Directive 2015/1535 defines an information society service as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services." 'Free' services, like advertisement-based services offered by social network companies, are included in the scope, as the Court of Justice of the European Union's (CJEU) Papasavvas and others decision (C-291/13) confirms.

According to the e-Commerce Directive, service providers are exempt from liability for illegal user-generated content if they expeditiously remove or disable access to the content upon obtaining knowledge or awareness of its unlawfulness (Art. 14). Member States shall not impose a general obligation on providers to monitor the information which they transmit or store, nor a general obligation to actively seek facts or circumstances indicating illegal activity (Art. 15). Nevertheless, the CJEU Eva Glawischnig-Piesczek v Facebook Ireland Limited decision (C-18/18) permits the national courts to oblige social networks to identify and delete comments identical to those previously deemed illegal. Critics of this decision warn of severe implications for the freedom of expression, since legal speech might get caught like 'dolphins in the net.'[133]

The European Commission submitted the Digital Services Act package consisting of the Digital Services Act (DSA) and the Digital Markets Act (DMA) to the European Parliament and the European Council in December 2020. The e-Commerce Directive will remain the basic legal framework and will only be updated and supplemented by the package. The package addresses technological trends like the spread

132 Council of Europe, 2021.
133 Keller, 2020.

of disinformation, exchange of illegal goods, online violence, privacy and targeted advertisement, etc., and represents an attempt to regulate the mounting power of technological giants by differentiating between hosting services, online platforms, and very large online platforms. The DMA deals with competition law aspects, while the DSA retains and updates the e-Commerce Directive's exemption from liability for service providers. According to the proposed DSA, every intermediary service provider will need to establish a point of contact for state authorities and a legal representative in the EU (Art. 10-13) and every hosting service provider will be obliged to provide mechanisms for flagging potentially illegal content and state the reasons for removal or blocking of content (Art. 14-15). There are additional obligations for online platforms to provide complaint-handling systems and dispute resolution, protection against illegal use of the platforms, as well as information obligations (Art. 17-24). Very large platforms will carry the additional obligations of security and control as well as more responsibilities regarding information and access (Art. 26-33). The DSA aims to make content moderation more transparent and force service providers to establish adequate redress procedures. The final shape and impacts of the proposed package remain to be seen, but critics warn that the proposal does not address social networks' 'opinion power' – that is, their political power.[134] Critics also describe it as both too ambitious and not ambitious enough, as its scope does not include 'harmful content' in general, but focuses on content that is illegal under EU or member state law.[135]

Slovenia first transposed the e-Commerce Directive by amending the Electronic Business and Electronic Signature Act (Zakon o elektronskem poslovanju in elektronskem podpisu – ZEPEP). In 2006, these provisions were transposed into the Electronic Commerce Market Act, which follows the EU definition of information society service and adopts a notice and takedown system when it comes to illegal user-generated content on Facebook and other social networks. Service providers are exempt from liability for user-generated content and are not obliged to monitor this content (Art. 8); however, they are required to stop and prevent violations by removing or blocking user-generated content when prompted by a court order (Art. 9-11). Once the social network is informed of the infringement, it ought to remove or block access to the illegal content 'expeditiously' (Art. 11). The exact meaning of the word 'expeditiously' is not defined. The variety of contexts implies diverse response times, thus it makes sense to establish the appropriate response time on a case-to-case basis.[136] If a service provider fails to act and such an omission results in damage, the provider may also face civil liability in accordance with Art. 131 of the Obligations Code (Obligacijski zakonik – OZ).

134 Helberger, 2020.
135 Morais Carvalho, Arga e Lima, and Farinha, 2021.
136 Damjan, 2017.

4.3. Slovenian legislation limiting the freedom of expression on social networks

4.3.1. Administrative law

Encouragement of intolerance is an administrative offense under the Protection of Public Order Act (Zakon o varstvu javnega reda in miru – ZJRM-1) and it may be punished by a fine (Art. 20). According to critics, this administrative offense is hardly distinguishable from the criminal offense of illegal hate speech examined below,[137] while others see it as a dangerous instrument of political power.[138] The Protection Against Discrimination Act (Zakon o varstvu pred diskriminacijo – ZVarD) establishes the Advocate of the Principle of Equality, an independent and autonomous state body mandated to deal with discrimination in both the public and private sector (Art. 1). The Act prohibits incitement of discrimination (Art. 10), without including incitement of discrimination among the administrative offenses (Art. 45). For example, when a breach of Art. 10 was detected in 2019 (case number 0700-53/2019), the Advocate issued an order establishing that the violation took place and demanded that discriminatory comments against the Roma community be removed from an online media's website. In the 2019 Annual Report, the Advocate urged that Art. 10 of the Protection Against Discrimination Act be requalified as an administrative offense and sanctioned with a fine.[139]

There is no administrative liability for the creation, dissemination, and usage of fake news in Slovenia. Media fitting the definition of the Mass Media Act are subject to the oversight of the Inspectorate for Culture and Media.[140] The Inspectorate is criticized as a *contradictio in adiecto* since the Mass Media Act does not prescribe the content that media ought to report.[141] The Inspectorate supervises compliance with the Act's provisions regarding the proper use of the Slovenian language, parental guidance advisories (e.g., I U 1228/2011), etc., but it does not supervise social media.

4.3.2. Criminal law

The chapter on criminal offenses against public order and peace in the Criminal Code (Kazenski zakonik – KZ-1) includes the prohibition of public incitement to hatred, violence, or intolerance (Art. 297). Such illegal hate speech includes public incitement of hatred, violence, or intolerance based on ethnicity, racial, religious or ethnic origin, sex, skin color, origin, wealth, education, social status, political or

137 Zobec, 2019.
138 Avbelj, 2018.
139 Lobnik, 2020.
140 "O inšpektoratu za kulturo in medije | GOV.SI," 2021.
141 Grčar, 2009.

other beliefs, disability, sexual orientation or any other personal circumstance; dissemination of ideas about the superiority of one race over another; and the denial, approval, justification, ridiculing, or advocating of genocide, the Holocaust, crimes against humanity, war crimes, aggression, or other crimes against humanity. The offense must be carried out in a manner likely to disturb public order or which is threatening, abusive, or insulting. The punishment of up to two years of imprisonment is foreseen, stretching to up to five years of imprisonment in cases of aggravated circumstances.

The State Prosecutor's Office of the Republic of Slovenia reported a slight increase in convictions for public incitement to hatred, violence, or intolerance in 2020 (six cases resulted in conviction).[142] This increase is partially due to a recent change in the interpretation of the hate speech provision. The criminal law definition of illegal hate speech was amended several times since Slovenia gained its independence in 1991. The latest of the amendments in 2011 restricted the applicability of Art. 297 to punish only conduct which is "either carried out in a manner likely to disturb public order or which is threatening, abusive or insulting," as permitted by Council Framework Decision 2008/913/JHA. Since this amendment, the Prosecutor's Office interpreted that Art. 297 is only applicable when the offense is carried out in a manner likely to disturb public order, ignoring the alternative dictum of the Council Framework and the Criminal Code.[143] Courts assumed the same interpretation – hate speech that was 'merely' threatening, abusive, or insulting was not considered a criminal offense.[144] This perception changed with the Supreme Court ruling that widened the scope of criminality to threatening, abusive, or insulting hate speech (I Ips 65803/2012). The case dealt with a public online post against the Roma community. While no concrete threat to public order was established, the Supreme Court declared the previous interpretation of Art. 297 erroneous and concluded that an abstract threat suffices to establish criminal liability.

Despite the global trend to decriminalize criminal offenses against honor and reputation or at least eliminate prison sentences for such offenses, the Criminal Code (Art. 158-165) threatens these sanctions for the following offenses: insult, slander, defamation, calumny, malicious false accusation of crime, insult to the Republic of Slovenia, insult to a foreign country or an international organization, and insult to the Slovenian people or national communities.[145] Most of these offenses (Art. 158-162) are prosecuted upon a private action (Art. 168) and are not sanctioned if the perpetrator was provoked or if they apologize or retract problematic statements (Art. 167). If offenses against honor and reputation are committed through the press, radio, television, or other means of public information or at a public assembly, or *on internet websites*, the threatened sanctions are more severe. The High Court of

142 "Skupno Poročilo o Delu Državnih Tožilstev Za Leto 2020," 2021.
143 Završnik, 2017; Ambrož, 2017.
144 Stajnko, Kičin, and Tomažič, 2020.
145 Korošec, Filipčič, and Zdolšek, 2018, pp. 842–847.

Ljubljana extended the circumstance 'on Internet websites' to the cases that transpired prior to the 2011 amendment that added it (II Kp 13079/2012). The Court held that the sanctions for the offenses against honor and reputation are generally rather mild but opined that a prison sentence may be consistent with the ECtHR interpretation of the freedom of expression, provided that the Court establishes that the context and gravity of the case demand it.

Offenses of unlawful publication of private writings (Art. 140) and abuse of personal information (Art. 143) prosecuted upon a complaint from the injured party or upon a private action, and the disclosure of classified information (Art. 260) prosecuted *ex offo*, can be committed by posting on social networks. The Criminal Code also includes a prohibition of incitement to violent change of the constitutional order (Art. 359). Social networks can play an important role in committing this offense, for instance, incitement to violent change of the constitutional order was committed by posting a video on a social network (XI Ips 40945/2018).

According to a 2015 study conducted by the Slovene Association of Journalists, out of 127 criminal charges against media reporting (i.e., against journalists, editors, and media houses), only six resulted in conviction; 43% of cases involved charges of defamation.[146] Defamation indicates that the truthfulness of media allegations can be established, unlike insult, which involves negative value judgments that do not have to be proven to be true or false (VII Kp 56216/2017).

4.3.3. Civil law

Civil law is the more popular avenue to seek legal protection and injunctive orders that may be used to demand a service provider's removal of an illegal post. The Obligations Code governs the request to cease the infringement of personal rights in Art. 134, prescribing the right to request that the court or any other relevant authority order that the infringement of the inviolability of the human person, personal and family life, or any other personal right be ceased, that such action be prevented, or that the consequences of such action be eliminated. Types of injunctive orders and the conditions that must be fulfilled for obtaining them are specified in the Enforcement and Security Act (Zakon o izvršbi in zavarovanju – ZIZ). If the infringement continues despite the court's order, monetary damages may follow. In the High Court of Ljubljana case I Cp 2892/2017, the defendant posted a range of personal data about the plaintiff on her Facebook profile, provoking numerous offensive and threatening comments. The Court found that immediate remedy is necessary and that waiting for a final judgment would render legal protection obsolete: Swift action was needed to prevent further damage.

According to the Obligations Code, the injured party may also demand a publication of judgement or correction (Art. 178) and/or monetary compensation (Art. 179). Defamation or calumny, assertion or dissemination of untrue statements on the

146 Delić and Stare, 2015.

past, knowledge, or capability of another resulting in material damage must be recompensed; liability is excluded if the speaker did not know that the information was untrue or if they had a genuine interest in so doing (Art. 177). Monetary compensation is also applicable in the case of physical or mental distress suffered owing to the defamation of good name or reputation, the curtailment of freedom or a personal right, and for fear caused, even if no material damage was inflicted (Art. 179). Since the Obligations Code does not define defamation of good name or reputation, civil courts utilize the definitions enshrined in the Criminal Code. In cases of offensive value judgments, a withdrawal of the statement may be ordered, while in the case of an offensive statement proven to be untrue, a revocation of the statement is an appropriate sanction (I Cp 2054/99). The court determines the amount of compensation for non-material damage based on the importance of the good affected and the purpose of the compensation if a causal link between the damage and the statements is proven.

The High Court of Koper dealt with a Facebook post expressing a warning against the plaintiff's brand of coffee (Cpg 213/2017). The Court opined that the language used was not offensive or depreciating and took the position that freedom of expression is a predisposition of existence for Facebook and that a public Facebook account does not necessarily mean that the post actually reaches all of the people active on Facebook. The Supreme Court stressed the importance of the freedom of (political) expression on social networks and prioritized it over the right to dignity, honor, reputation, and personal dignity (II Ips 75/2019). While the Twitter post in question was vulgar, the Court warned against the chilling effect of sanctioning such speech. The Court regarded the context of the Twitter social network: the specific style and manner of speech qualified by short, fast, and vulgar communication that is produced spontaneously and consumed quickly and without much reflection by the ordinary user. The court considered the defendant's Twitter profile's large following but decided that the plaintiff, a public personality, should tolerate more nuisance than an ordinary citizen. The High Court of Ljubljana produced a different understanding of political critique expressed on social networks, finding that the harsh language used by the defendant was not a political critique but rather an attempt to depreciate the local major by comparing him to Hitler (II Cp 701/2015). The plaintiff's right to honor, reputation, and personal dignity were prioritized over the freedom of expression. The High Court of Ljubljana case II Cp 577/2019 involved satirical publications in a closed Facebook group with 67 members. The Court took into consideration that the group involved only people belonging to the local community. It also underlined the specific nature of political satire that permits a wide range of expression and found no violation. In the High Court of Ljubljana case II Cp 2066/2012, a photograph taken in the plaintiff's home was published on a Facebook profile set to private. The Court asserted that although the post was meant for the 'closest friends,' the photograph was made public and commented upon, with some of the comments being offensive to the plaintiff, and thus constituted a breach of his privacy. In the High Court of Maribor case I Cp 193/2012, the Court concluded that a student filming an extraordinary event at a

public gathering is something the plaintiff should have expected and did not find his posting of the video on Facebook illegal.

Case law reveals that Slovenian courts balance the personality rights of one party against the freedom of expression of another. The balance is difficult to establish and must take into consideration the characteristics of each individual case: The nature of the social network, the context of the post, the number of people with access to the post, and the expectation that certain acts might be photographed, recorded, and posted online are all taken into account.

The Mass Media Act grants the right to correction to anyone who feels offended or insulted by media reporting (Art. 26-41). The insulted party may demand a publication of a correction that has to be published in the same way as the original content. If the correction is valid (Art. 31), the editor must publish it within 24 hours or risk a civil lawsuit. The correction does not need to be true or correct, since the essence of this right is not in establishing the truth but in enabling the person who felt injured by the reporting to respond (II Cp 2634/2017). To ensure objective, plural, and timely information, anyone has the right to demand the publication of a verifiable response denying, correcting, or supplementing reported information (Art. 42-44). If the editor judges that the response is valid, it must be published. Social media platforms need not guarantee the rights to correction and response (technically though, anyone may open an account and respond to the post). Publishing misinformation may result in damages if statements are based on facts there were not properly researched and checked before publication (II Cp 1666/2014).

4.3.4. Journalists' professional liability

Journalists self-regulate the ethical aspects of their profession with codes of conduct. Like the legal framework, self-regulation is lagging behind the developments of technology ('Google reporting' – fact-checking using nothing but an Internet browser, citizen reporters and bloggers, social media, etc.).[147] The Slovenian Union of Journalists and the Slovene Association of Journalists are the most important actors in the field of journalists' self-regulation.[148] Violations of the Code of Journalists' Conduct are subject to the Journalists' Court of Honor, which is composed of nine journalists and two representatives of the public.[149] The key ethical principles in the Code are freedom of expression, verifying information, and avoiding causing harm to those reported about. In 2019, the Journalist Court received 47 complaints and has established violations of the Code in 47% of the cases.[150] Alleged dissemination of mis- and disinformation is the most common ground for complaints: 23 of the 47 complaints in 2019 were referring to the lack of fact-checking and due diligence

147 Kovačič, 2014.
148 "Sindikat novinarjev Slovenije," 2021; Društvo novinarjev Slovenije, 2021.
149 "Člani NČR – Novinarsko častno razsodišče," 2021.
150 Stare, 2020.

in preventing the spread of misinformation; violations were found in six of these cases.[151] The Journalists' Court of Honor decides cases involving professional and amateur journalists. It cannot punish journalists, but it may suggest the expulsion of journalists from their media organizations.[152]

The public media house RTV Slovenija is regulated by the Radiotelevizija Slovenija Act (Zakon o Radioteleviziji Slovenija – ZRTVS-1) and it functions as a non-profit organization of special cultural and national importance, providing a wide range of informative, educational, cultural, and other content to serve its audience. The journalists of RTV Slovenija have their own professional code of conduct.[153] The code establishes the Guardian of Professional and Ethical Standards that acts in the interest of all concerned parties. The Guardian is the point of reference for complaints and suggestions and is responsible to act upon them and report the results to the involved parties. The Guardian publishes regular reports available to the general public.

Project Oštro, a center for investigative journalism in the Adriatic region, is another attempt from journalists to respond to the increasing levels of mis- and disinformation in the media.[154] Covering Slovenia, Croatia, and Italy, Oštro's investigative journalists fact-check media stories and respond to misleading information. Oštro has its own code of conduct that is based on the values of independence, non-profit activity, and democratic debate.[155] The European project Open Your Eyes is another attempt to offer reliable information by establishing a database that can be used as a tool when discriminating between information and disinformation.[156] Several other projects studying and countering the coronavirus 'infodemic' and disinformation-related knowledge are funded by the European Commission.[157] In line with the Code of Practice on Disinformation, social media are also removing mis- and disinformation from their platforms.[158]

5. The impact of content moderation on freedom of expression and pluralism

Freedom of expression is a prerequisite for an open democratic society, yet no freedom is absolute. The 1990s was a decade intoxicated by the idea that the Internet will enable large-scale participation in debates in such a way as to escape the

151 Ibid.
152 "Novinarsko častno razsodišče – Novinarsko častno razsodišče," 2021.
153 "Pravilnik o poklicnih standardih," 2021.
154 "O Oštru: Center za preiskovalno novinarstvo v jadranski regiji," 2019.
155 "Oštrov kodeks," 2020.
156 "Fake News for Dummies: Check It Out," 2021.
157 European Commission, 2021d.
158 European Commission, 2021b.

traditional means of control, strengthen and reinforce the democratic structures, permit global cooperation, and allow people to self-regulate diverse cyberspaces.[159] In practice, the Internet is dominated by the profit-driven enterprises of multinational corporations whose algorithms contribute to the rise of incendiary speech and mis- and disinformation. More speech does not necessarily mean better speech – hate speech, threats, and insults have often been used to silence certain groups and might stiffen the pluralism of the public debates.[160] The same trends are emerging in Slovenia: While social media allow anyone to express and circulate their ideas in principle, vulgar and offensive language often trumps nuanced discussion.[161] People find themselves targeted and silenced by anonymous users,[162] which contributes to the polarization of society and leaves freedom of expression up for grabs, that is, available to the loudest and most aggressive speakers. The idea of democratic debate, in contrast, presupposes a minimal level of civility and the use of arguments. To ensure open participation in democratic debate, the Slovenian legal system restricts individuals' freedom of expression in balance with the freedoms and rights of others. While social networks are scarcely regulated, they must remove illegal speech from their platforms. When it comes to Facebook's own content moderation, Slovenian users are left to follow Facebook's appeal process if they feel that their expression was limited without grounds. Unproblematic content sometimes gets removed and this might have negative implications for freedom of expression and pluralism on the platform – more transparency and democratic accountability of social networks would surely improve the situation. It will be interesting to see how the introduction of the Oversight Board and the DSA will impact these issues that will certainly generate more controversy in the future.

The media sphere in Slovenia is deeply marked by the process of economic and political transition[163] and is vulnerable to the interests of politics and capital; many Slovenian reporters find themselves in a position of precarious labor relations that increase self-censorship.[164] Simultaneously, widespread Internet use is changing journalism, which is increasingly perceived as a practice in which anyone can engage; people expect to consume news for free and speed is often prioritized over due investigation.[165] Reports on political leaders' intimidation and attrition of journalists and dwindling freedom of press in Slovenia are growing.[166] The Slovene Association of Journalists has also reported a rise in the abuse of legal remedies for the financial and mental attrition of journalists – a practice known as a strategic lawsuit against

159 Johnson and Post, 1996.
160 Keats Citron, 2019.
161 Jereb, 2020.
162 Voh Boštic, 2021.
163 Hrvatin and Petković, 2007.
164 Čeferin, Poler, and Milosavljević, 2017.
165 Črnič, 2007.
166 Ombudsman RS, 2021; European Parliament, 2021; Reporters Without Borders, 2021; Wiseman, 2020; Bayer, 2021.

public participation (SLAPP).[167] In this context, social media represent a new frontier in the way news is created, distributed, and consumed: More and more people rely on social networks as a news source. Social media contribute to the pluralization of the public debate and enable the active participation of people traditionally perceived as mere consumers of news.[168] On the other hand, the rise of social media is accompanied by the spread of mis- and disinformation. The algorithms controlling the content social media users see follow users' preferences intuited from a vast array of data about each user: Users are shown the news that is more likely to grab their attention and conform to their beliefs, contributing to the so-called 'filter-bubbles' or 'echo chambers' where people are only exposed to narratives they are likely to agree with, making them vulnerable to extreme polarization.[169] Furthermore, fake news tends to sound interesting and usually receives more clicks and shares than other content.[170] Users of social media do not seem to engage with the news very profoundly, and news stories are often shared without being read, causing mis- and disinformation to spread faster than actual news.[171]

Legal scholars warn of several practical and legal problems related to the idea of sanctioning the creation and dissemination of fake news: The definition of fake news is too open-ended, the deceitful intent is difficult to prove, perpetrators are often anonymous, and democratic values like freedom of expression are at stake. On the other hand, the spread of disinformation can have several displeasing consequences that beg for a regulatory response. The Covid-19 outbreak has motivated several states to tighten the rules about the spread of misinformation, most notoriously Russia, which supplemented its list of administrative fake news offenses with the criminal offense of the deliberate spread of false information about serious matters of public safety.[172] The Russian amendment of the Criminal Code is widely perceived as another attempt to crack down on government critics.[173] Malaysia also recently adopted an anti-fake news decree that is seen as the government's attempt to impose its own version of the truth.[174] The EU, on the other hand, favors self-regulation.

Germany, the pioneer of the self-regulatory approach, eventually judged it inefficient and responded with the Network Enforcement Act – Netzwerkdurchsetzungsgesetz (NetzDZ) in 2018. The German solution of binding social media to remove obviously unlawful content within 24 hours has been widely criticized for inciting private censorship, operating with vague notions, and clashing with the EU legislation.[175] France has created a new civil procedure to prevent the transmission

167 Društvo novinarjev Slovenije, 2020.
168 Noor, 2016; Črnič, 2007.
169 Pariser, 2012; Napoli, 2018.
170 Fallis and Mathiesen, 2019.
171 Gabielkov et al., 2016.
172 International Press Institute, 2020.
173 Ibid.
174 Reporters Without Borders, 2021.
175 Claussen, 2018.

of factually inaccurate or misleading information in order to protect public order and the integrity of elections with a recently adopted German-style law (Loi du 24 Juin 2020 Visant à Lutter Contre les Contenus Haineux sur Internet).[176] The common thread linking these attempts is to assign social media companies more responsibility to monitor and moderate user-generated content. The threat of fines is a strong incentive to remove suspicious or reported content, even when unproblematic. None of the national legislative initiatives mentioned engage with the algorithmic architecture behind the phenomenon of fake news. However, filter bubbles and clickbait are the heart of the social media business model. Perhaps more attention could be paid to social media companies' manipulative practices, forcing the companies to be more transparent and offer users more control over the process that determines what kind of content they are shown. While the overall problems of online hate speech and mis- and disinformation can probably never be efficiently solved by legislative measures alone, this might be an important step toward making tech companies responsible not only for the content users see and share on their platforms, but also for the business practices that determine how, why, and which content is shown to a specific user. A transnational approach would be the most appropriate, as the problems are global.

6. Conclusions

Since social media host user-generated content and do not create or edit the content on their platforms, they are not considered to be 'media' under the jurisdiction of the Slovenian Mass Media Act. Imposing more obligations on social networks by recognizing them as a type of media organization would increase their liability for the content they host and may define their obligation to both guarantee and limit users' freedom of expression. As host service providers under the e-Commerce Directive and the Slovenian Electronic Commerce Market Act, they are excluded from liability for the hosted content if they are not aware that such content is illegal and must only remove it once its illegality is established. While mis- and disinformation on social media influence people's perceptions of reality and contribute to the general cynicism, it is extremely difficult to efficiently regulate it. It is impossible to conclusively draw the fine line between necessary regulation and freedom of expression, between an opinion and a lie, and so on.

While social networks get to decide which content they do not wish to host, they can be legally forced to remove illegal content. An overview of Slovenian criminal and civil case law involving various types of social network posts reveals that courts operate under the assumption that freedom of expression on social networks must be

176 Helberger, 2020.

protected and accept that the standards of expression on these networks are lower compared to other social contexts. Case law clearly demonstrates that privacy expectations have been transformed in the digital age. The number of social media followers or group members is also an important factor in court decisions, yet its interpretation largely depends on the other facts of a case. Courts decide not to take concepts like 'private' or 'public' post/group/account at face value; rather, they strive to understand the full context and personal circumstances of those involved. A large number of court cases dealing with social network posts confirms that social networks play a very important and sometimes controversial role in people's daily lives, while the diverse argumentations and decisions taken by Slovenian courts testify that social networks are nevertheless a relatively new phenomenon and a small piece of the puzzle constituting each individual case.

While it is difficult to believe that the Slovenian state would do a better job than Facebook at regulating speech on the platform, Facebook should not be perceived as benevolent, capable, or an appropriate entity to decide on the (il)legality of user-generated content. Procedures, oversight, and legal remedies for users should be made available. However, what is needed the most is a more thorough reform based on cooperation between states, companies, and the citizens of the globe. Furthermore, users should be more adequately informed about these companies' modus operandi (the algorithmic architecture of their platforms, data use, advertising practices, options to opt out, rules and procedures for content moderation available in the local language, etc.), have a say in the rules and procedures, and be recompensed for their data and time. The EU has an important role in this process and is attempting to address these very issues and curb tech companies' power with the proposed Digital Services Act package whose final shape and effects remain to be seen. In the final instance, the issue of freedom of expression on social networks is global and it demands transnational regulatory responses.

Bibliography

Abram, A., Grušovnik, T. (2021) *Koronavirus in teorije zarot* [Online]. Available at https://znc.si/dogodki/znanost-med-knjigami/koronavirus-in-teorije-zarot/. (Accessed: 2 April 2021).

Ambrož, M. (2017) 'Sovražni govor kot ogrozitveno kaznivo dejanje', in: Splichal, S. ed. *Zagovor javnosti: Med svobodo izražanja in sovražnim govorom*. Ljubljana, Slovenska akademija znanosti in umetnosti, pp. 90–100.

Anon (2010) *Srb v Šarm el Šejku ubil smrtonosnega morskega psa* [Online]. Available at https://siol.net/novice/novice/srb-v-sarm-el-sejku-ubil-smrtonosnega-morskega-psa-57887. (Accessed: 6 April 2021)

Anon (2018) *Protest proti zakonu, ki bi prepovedal vpis necepljenih otrok v javne vrtce* [Online]. Available at https://www.rtvslo.si/zdravje/protest-proti-zakonu-ki-bi-prepovedal-vpis-necepljenih-otrok-v-javne-vrtce/452120. (Accessed: 2 April 2021)

Anon (2019) *O Oštru: Center za preiskovalno novinarstvo v jadranski regiji* [Online]. Available at https://www.ostro.si/si/center. (Accessed: 9 April 2021)

Anon (2020) *Janša: Lažne novice se v Sloveniji širijo tudi v osrednjih medijih* [Online]. Available at http://www.times.si/slovenija/jansa-lazne-novice-se-v-sloveniji-sirijo-tudi-v-osrednjih-medijih--b34640fa00998b12b9395e44be8320feb1db009b.html. (Accessed: 21 April 2021)

Anon (2020) *Mit o cepivu razkrit, ni nevarnosti za neplodnost* [Online]. Available at https://www.delo.si/razno/mit-o-cepivu-razkrit-ni-nevarnosti-za-neplodnost/. (Accessed: 2 April 2021)

Anon (2020) *Oštrov kodeks* [Online]. Available at https://www.ostro.si/si/kodeks. (Accessed: 9 April 2021)

Anon (2020) *Strokovnjaki: Cepivo proti covidu-19 ne povzroča poškodb posteljice* [Online]. Available at https://www.rtvslo.si/zdravje/novi-koronavirus/strokovnjaki-cepivo-proti-covidu-19-ne-povzroca-poskodb-posteljice/544837. (Accessed: 2 April 2021)

Anon (2021) *Astroturfing* [Online]. Available at https://isjfr.zrc-sazu.si/sl/terminologisce/svetovanje/astroturfing. (Accessed: 6 April 2021)

Anon (2021) *Koalicija proti sovražnim napadom. Opozicija: Tudi sami morate prevzeti odgovornost.* [Online]. Available at https://www.rtvslo.si/slovenija/koalicija-proti-sovraznim-napadom-opozicija-tudi-sami-morate-prevzeti-odgovornost/570210. (Accessed: 10 March 2021)

Anon (2021) Referring Former President Trump's Suspension From Facebook to the Oversight Board. Available at https://about.fb.com/news/2021/01/referring-trump-suspension-to-oversight-board/. (Accessed: 8 March 2021)

Anon (2021) *Skupno poročilo o delu državnih tožilstev za leto 2020*. Ljubljana, Public Prosecutor's Office of Republic of Slovenia.

Anon (2021) *V Mariboru večstoglava množica vzklikala: 'Maske dol, vlada pa v zapor'* [Online]. Available at https://www.rtvslo.si/lokalne-novice/maribor/v-mariboru-vecstoglava-mnozica-vzklikala-maske-dol-vlada-pa-v-zapor/575438. (Accessed: 6 April 2021)

Anon (n.d.) *Bylaws – Oversight Board* [Online]. Available at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjs3cirg6vvAhXkwAIHHYjhAR0QFjAAegQIARAD&url=https%3A%2F%2Foversightboard.com%2Fsr%2Fgovernance%2Fbylaws&usg=AOvVaw1d35jmY-5eTVbF4-YfwC0D. (Accessed: 3 December 2021)

Anon (n.d.) Člani NČR – Novinarsko častno razsodišče. Available at https://razsodisce.org/o-ncr/clani-ncr/. (Accessed: 9 April 2021)

Anon (n.d.) *Community Standards | Facebook* [Online]. Available at https://www.facebook.com/communitystandards/. (Accessed: 18 February 2021)

Anon (n.d.) *Community Standards Enforcement* [Online]. Available at https://transparency.facebook.com/community-standards-enforcement/guide. (Accessed: 5 March 2021)

Anon (n.d.) *Družabna omrežja | safe.si* [Online]. Available at https://safe.si/nasveti/druzabna-omrezja. (Accessed: 26 March 2021)

Anon (n.d.) *Facebook: Terms of Service* [Online]. Available at https://www.facebook.com/terms.php. (Accessed: 1 April 2021)

Anon (n.d.) *Fake news for dummies: baza podatkov Check It Out* [Online]. Available at https://openyoureyes.info/si/database/list. (Accessed: 9 April 2021)

Anon (n.d.) *I don't think Facebook should have taken down my post. | Facebook Help Center* [Online]. Available at https://www.facebook.com/help/2090856331203011. (Accessed: 23 February 2021)

Anon (n.d.) *Kampanja Ne sovražnemu govoru | GOV.SI* [Online]. Available at https://www.gov.si/zbirke/projekti-in-programi/kampanja-ne-sovraznemu-govoru/. (Accessed: 10 March 2021)

Anon (n.d.) *Maske Dol | Facebook* [Online]. Available at https://www.facebook.com/rizzka.chiyuzschiyuzs. (Accessed: 6 April 2021)

Anon (n.d.) *Most used social media 2020* [Online]. Available at https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/. (Accessed: 19 February 2021)

Anon (n.d.) *My Personal Account Was Disabled | Facebook* [Online]. Available at https://www.facebook.com/help/contact/260749603972907. (Accessed: 23 February 2021)

Anon (Facebook groups) *Nasprotniki protestirali proti obveznemu cepljenju. Strokovnjaki: ker določenih bolezni ne vidimo, se ne zavedamo nevarnosti* [Online]. Available at https://www.24ur.com/novice/slovenija/nasprotniki-cepljenja-na-protestu-moja-hcerka-je-zbolela-za-mozganskim-encefalitisom-zaradi-cepiva.html. (Accessed: 2 April 2021)

Anon (n.d.) Novinarsko častno razsodišče – Novinarsko častno razsodišče. Available at https://razsodisce.org/o-ncr/novinarsko-castno-razsodisce-organ-dns-in-sns/. (Accessed: 9 April 2021)

Anon (n.d.) *O inšpektoratu za kulturo in medije | GOV.SI* [Online]. Available at https://www.gov.si/drzavni-organi/organi-v-sestavi/inspektorat-za-kulturo-in-medije/o-inspektoratu/. (Accessed: 9 April 2021)

Anon (n.d.) *Oversight Board | Independent Judgment. Transparency. Legitimacy.* [Online]. Available at https://www.oversightboard.com/. (Accessed: 8 March 2021)

Anon (n.d.) *Oxford Word of the Year 2016 | Oxford Languages* [Online]. Available at https://languages.oup.com/word-of-the-year/2016/. (Accessed: 18 March 2021)

Anon (n.d.) *Pravilnik o poklicnih standardih* [Online]. Available at https://www.rtvslo.si/rtv/kdo-smo/zakoni-predpisi-in-dokumenti/pravilnik-o-poklicnih-standardih/475078. (Accessed: 9 April 2021)

Anon (n.d.) *r/Slovenia* [Online]. Available at https://www.reddit.com/r/Slovenia/.

Anon (n.d.) Sindikat novinarjev Slovenije. Available at http://sindikat-novinarjev.si/. (Accessed: 9 April 2021)

Anon (n.d.) *Skupaj za zdravje človeka in narave | Facebook* [Online]. Available at https://www.facebook.com/SkupajZaZdravje. (Accessed: 2 April 2021)

Anon (n.d.) *Sovražni govor na spletnih družbenih omrežjih v Sloveniji* [Online]. Available at https://www.mirovni-institut.si/projekti/sovrazni-govor-na-spletnih-druzbenih-omrezjih-v-sloveniji/. (Accessed: 12 March 2021)

Anon (n.d.) *The Facebook Company Products | Facebook Help Center* [Online]. Available at https://www.facebook.com/help/195227921252400. (Accessed: 19 February 2021].

Anon (n.d.) *Trditve in dejstva o Spletnem očesu | Spletno oko* [Online]. Available at https://www.spletno-oko.si/novice/trditve-in-dejstva-o-spletnem-ocesu. (Accessed: 23 February 2021)

Anon (n.d.) *Z (od)govorom na sovražni govor – ZaGovor* [Online]. Available at https://www.mirovni-institut.si/govor/. (Accessed: 10 March 2021)

Anon (n.d.) *Združenje za naravni razvoj otrok | Facebook* [Online]. Available at https://www.facebook.com/naravnirazvojotrok. (Accessed: 2 April 2021)

Avbelj, M. (2018) *Pravljičnih 7 o sovražnem govoru* [Online]. Available at https://www.iusinfo.si/medijsko-sredisce/kolumne/233434. (Accessed: 20 April 2021)

Bajt, V. (2017a) 'Online Hate speech and the "refugee crisis" in Slovenia', in: Kogovšek Šalamon, Ž., Žagar, I.Ž. and Lukšič-Hacin, M. (eds.) *The disaster of European refugee policy: Perspectives from the 'Balkan route'*. pp. 133–155.

Bajt, V. (2017b) 'Sovražni govor kot spodbuda kritičnega delovanja', in: Splichal, S. ed. *Zagovor javnosti: Med sovobodo izražanja in sovržanim govorom*. Ljubljana, Slovenska akademija znanosti in umetnosti, pp. 70–77.

Bakir, V., McStay, A. (2018) 'Fake News and The Economy of Emotions', *Digital Journalism*, 6 (2), pp. 154–175.

Balkin, J. (2014) 'Old-School/New-School Speech Regulation', *Faculty Scholarship Series* [Online], Available at https://digitalcommons.law.yale.edu/fss_papers/4877.

Balkin, J. (2018) 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', *Faculty Scholarship Series* [Online], Available at https://digitalcommons.law.yale.edu/fss_papers/5160.

Bašić Hrvatin, S. (2020) 'Konec medijskih tovarn', in: Lazar, I., Panjek, A. and Vinkler, J. (eds.) *Mikro in makro: Pristopi in prispevki k humanističnim vedam ob dvajsetletnici UP Fakultete za humanistične študije (2. knjiga)*. Koper, Založba Univerze na Primorskem, pp. 643–658.

Bates, D. (2010) *Shark terrorising Egyptian resort 'killed by drunken Serb who jumped on its head': The hoax taking the internet by storm* [Online]. Available at https://www.dailymail.co.uk/news/article-1340624/Egypt-shark-killed-drunken-Serb-jumped-head.html. (Accessed: 6 April 2021)

Bayer, L. (2021) *Inside Slovenia's war on the media* [Online]. Available at https://www.politico.eu/article/slovenia-war-on-media-janez-jansa/. (Accessed: 7 April 2021)

Bossaller, J., Bernier, A., McQueen, S., Peterson, M. (2019) The hoax and the president: Historical perspectives on politics, truth, and academia. Available at https://www.ideals.illinois.edu/handle/2142/105351. (Accessed: 18 March 2021)

boyd, danah m., Ellison, N. B. (2007) 'Social Network Sites: Definition, History, and Scholarship', *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230.

Bricket, M. (2018) Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process. Available at https://about.fb.com/news/2018/04/comprehensive-community-standards/. (Accessed: 18 February 2021)

Brügger, N. (2015) 'A brief history of Facebook as a media text: The development of an empty structure', *First Monday*, 20(5).

Burkhardt, J. M. (2017) History of Fake News. *Library Technology Reports*, 53(8), pp. 5–9.

Čeferin, R., Poler, M., Milosavljević, M. (2017) 'Prekarno delo novinarjev kot grožnja svobodi izražanja: odgovornost državnih oblasti', *Javnost – The Public*, 24(sup1), pp. 47–63.

Chander, A. (2012) 'Facebookistan', *North Carolina Law Review*, 90(5), p. 1807.

Chesney, R., Citron, D. (2019) 'Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics Essays', *Foreign Affairs*, 98(1), pp. 147–155.

Chesney, R., Citron, D. K. (2018) *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* [Online]. Report no. ID 3213954. Rochester, NY, Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3213954. (Accessed: 18 March 2021)

Cijan, N. (2020) *Ponarejeni videoposnetki (deepfake video)* [Online]. Available at: https://www.rtvslo.si/kultura/film-in-tv/ponarejeni-videoposnetki-deepfake-video/511782. (Accessed: 19 March 2021)

Claussen, V. (2018) 'Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation', *Media Laws*, (3).

Council of Europe (2020) *Guide on Article 10 of the European Convention on Human Rights: Freedom of expression*. European Court of Human Rights.

Council of Europe (n.d.) *Internet Intermediaries* [Online]. Available at: https://www.coe.int/en/web/freedom-expression/internet-intermediaries. (Accessed: 25 February 2021)

Črnič, T. O. (2007) 'Spletno novinarstvo skozi optiko novinarjev', *Družboslovne razprave*, 23 (54), pp. 43–64.

Damjan, M. (2017) 'Odgovornost ponudnikov storitev informacijske družbe in spletnih medijev za uporabniške vsebine', *Pravosodni bilten*, (1), p. 153.

Damjan, M. (2019) 'New Regulation of Copyright and Related Rights in the Digital Single Market', *Pravnik: Revija za Pravno Teorijo in Prakso*, 74(3–4), pp. 169–176.

Delić, A. and Stare, Š. (2015) Analiza tožb in ovadb medijev. Available at: https://novinar.com/novica/analiza-tozb-in-ovadb-medijev/. (Accessed: 13th April 2021)

Društvo novinarjev Slovenije (2020) Zloraba tožb za zastraševanje ter finančno in administrativno izčrpavanje medijev? Available at: https://novinar.com/novica/zloraba-tozb-za-zastrasevanje-ter-financno-in-administrativno-izcrpavanje-medijev/. (Accessed: 13th April 2021)

Društvo novinarjev Slovenije (n.d.) Kodeks. Available at: https://novinar.com/drustvo-novinarjev-slovenije/o-nas/dokumenti/kodeks/. (Accessed: 9 April 2021)

European Commission (2021a) *Code of Practice on Disinformation* [Online]. Available at: https://ec.europa.eu/digital-single-market/en/code-practice-disinformation. (Accessed: 18 March 2021)

European Commission (2021b) *Coronavirus disinformation: extended platforms' monitoring programme with focus on vaccines | Shaping Europe's digital future* [Online]. Available at: https://digital-strategy.ec.europa.eu/en/news/coronavirus-disinformation-extended-platforms-monitoring-programme-focus-vaccines. (Accessed: 21 April 2021)

European Commission (n. d.) *Facebook changes its terms and clarify its use of data for consumers* [Online]. Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_2048. (Accessed: 23 March 2021a)

European Commission (n. d.) *Funded projects in the fight against disinformation* [Online]. Available at: https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response-copy/fighting-disinformation-old/funded-projects-fight-against-disinformation_en. (Accessed: 21 April 2021b)

European Commission (n. d.) *Identifying conspiracy theories* [Online]. Available at: https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation/identifying-conspiracy-theories_en. (Accessed: 2 April 2021c)

European Commission (n. d.) *The EU Code of conduct on countering illegal hate speech online* [Online]. Available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en. (Accessed: 1 March 2021d)

European Parliament (2021) *Media freedom in Slovenia: MEPs to continue taking stock of recent developments | News | European Parliament* [Online]. Available at: https://www.europarl.europa.eu/news/en/press-room/20210322IPR00514/media-freedom-in-slovenia-meps-to-continue-taking-stock-of-recent-developments. (Accessed: 7 April 2021)

Facebook (2021) *Corporate Human Rights Policy*.

Facebook (n. d.) *Standardi skupnosti | Facebook* [Online]. Available at: https://sl-si.facebook.com/communitystandards/. (Accessed: 16 February 2021a)

Facebook (n. d.) Available at: https://m.facebook.com/communitystandards/hate_speech/. (Accessed: 13th July 2021b)

Fallis, D., Mathiesen, K. (2019) 'Fake news is counterfeit news', *Inquiry*, 0 (0), pp. 1–20.

Fišer, D., Erjavec, T., Ljubešić, N. (2016) 'JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin', *Slovenščina 2.0: empirical, applied and interdisciplinary research*, 4 (2), pp. 67–99.

Gabielkov, M., Ramachandran, A., Chaintreau, A., Legout, A. (2016) 'Social Clicks: What and Who Gets Read on Twitter? In: *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, SIGMETRICS '16. New York, NY, USA, Association for Computing Machinery, pp. 179–192.

Galvan, B. (2020) 'Facebook's Legal Responsibility for the Rohingya Genocide Comments', *University of San Francisco Law Review*, 55 (1), pp. 123–152.

Germani, F., Biller-Andorno, N. (2021) 'The anti-vaccination infodemic on social media: A behavioral analysis', *PLOS ONE*, 16 (3).

Gillespie, T. (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.

Godnov, U., Redek, T. (2014) 'The use of Twitter for political purposes in Slovenia', *Romanian Journal of Political Science*, 14 (1).

Gorenšek, T. (2019) *Izziv prihajajočega desetletja: 'deepfake' tehnologija* [Online]. Available at: https://www.24ur.com/novice/slovenija/izziv-prihajajocega-desetletja-deepfake-tehnologija.html. (Accessed: 19 March 2021)

Grčar, C. (2009) 'Nekateri praktični problemi medijskega prava', in: Lampe, R. ed. *Medijsko pravo*. Ljubljana, Planet GV, pp. 140–154.

Greenfield, H. (2018) EU Countries' Adopted Position on Copyright Will Hurt Internet Users and Digital Sector. Available at: https://www.ccianet.org/2018/05/eu-countries-adopted-position-on-copyright-will-hurt-internet-users-and-digital-sector/. (Accessed: 23 March 2021)

Grimmelmann, J. (2015) 'The Virtues of Moderation', *Yale Journal of Law and Technology*, 17 (1).

Hannah, M. (2021) QAnon and the information dark age. *First Monday* [Online], Available at: https://firstmonday.org/ojs/index.php/fm/article/view/10868. (Accessed: 2 April 2021)

Harris, B. (2020a) Oversight Board to Start Hearing Cases. Available at: https://about.fb.com/news/2020/10/oversight-board-to-start-hearing-cases/. (Accessed: 18 February 2021)

HARRIS, B. (2020b) Preparing the Way Forward for Facebook's Oversight Board. Available at: https://about.fb.com/news/2020/01/facebooks-oversight-board/. (Accessed: 9 March 2021)

HARRIS, D. (2018) 'Deepfakes: False Pornography Is Here and the Law Cannot Protect You', *Duke Law & Technology Review*, 17, pp. 99–128.

HEINS, M. (2013) 'The Brave New World of Social Media Censorship', *Harvard Law Review Forum*, 127, pp. 325–330.

HELBERGER, N. (2020) 'The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power', *Digital Journalism*, 8(6), pp. 842–854.

HORVAT, U. (2020) Slovenia as the last EU country that does not yet have a GDPR implementing law. Available at: http://jadek-pensa.si/en/slovenia-as-the-last-eu-country-that-does-not-yet-have-a-gdpr-implementing-law/. (Accessed: 29 March 2021)

HRVATIN, S. B., PETKOVIĆ, B. (2007) *In temu pravite medijski trg? Vloga države v medijskem sektorju v Sloveniji*. Ljubljana, Mirovni inštitut.

HUNT, J. (2017) *Fake news is officially 2017's word of the year* [Online]. Available at: https://www.independent.co.uk/news/uk/home-news/fake-news-word-year-2017-collins-dictionary-donald-trump-kellyanne-conway-antifa-corbynmania-gender-fluidity-fidget-spinner-a8032751.html (Accessed: 18 March 2021).

HUTCHINSON, A. (2020) *New Fake Account Removals Highlight Twitter's Bot Problem Once Again* [Online]. Available at: https://www.socialmediatoday.com/news/new-fake-account-removals-highlight-twitters-bot-problem-once-again/575488/. (Accessed: 6 April 2021)

IBRAHIM, Y. (2017) 'Facebook and the Napalm Girl: Reframing the Iconic as Pornographic', *Social Media + Society*, 3 (4).

International Press Institute (2020) 'New 'fake news' law stifles independent reporting in Russia on COVID-19. Available at: https://ipi.media/new-fake-news-law-stifles-independent-reporting-in-russia-on-covid-19/. (Accessed: 20th April 2021)

JALUŠIČ, V. (2019) 'Ko gre jezik na dopust', *Časopis za kritiko znanosti*, (268), pp. 21–42.

JANČIČ, P. (2017) *Fake news – Lažnive novice*. Ljubljana, Družina.

JENKO, P. (2018) *Manipulacija videoposnetkov: Deepfake posnetki kot znanilci nove dobe lažnih novic* [Online]. Available at: https://www.dnevnik.si/1042819488. (Accessed: 19 March 2021)

JEREB, A. (2020) 'Žaljive objave na družbenih omrežjih kot moderni izziv sodišč: so lahko tudi všečki protipravni?', *Pravni letopis 2020*, pp. 175–199.

JOHNSON, B. J. (2016) 'Facebook's Free Speech Balancing Act: Corporate Social Responsibility and Norms of Online Discourse', *University of Baltimore Journal of Media Law & Ethics*, 5 (3–4), pp. 19–37.

JOHNSON, D. R., POST, D. (1996) 'Law and Borders: The Rise of Law in Cyberspace', *Stanford Law Review*, 48(5), pp. 1367–1402.

JONTES, D. (2010) *Novinarstvo kot kultura: Miti in vrednoste*. Ljubljana, Založba FDV.

JOVANOVIĆ, M. (2020) *'Njuzovci' na tribini u Nišu: Sve nam je teže, konkurencija u objavljivanju lažnih vesti sve veća* [Online]. Available at: https://www.juznevesti.com/Drushtvo/Njuzovci-na-tribini-u-Nisu-Sve-nam-je-teze-kokurencija-u-objavljivanju-laznih-vesti-sve-veca.sr.html. (Accessed: 6 April 2021)

JUKOVIČ, G. (2017) *STAznanost: Lažne novice so produkt družbe, ki išče preproste rešitve za kompleksne probleme (II)* [Online]. Available at: http://znanost.sta.si/2350199/lazne-novice-so-produkt-druzbe-ki-isce-preproste-resitve-za-kompleksne-probleme-ii. (Accessed: 17 March 2021)

KADRI, T., KLONICK, K. (2019) 'Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech', *Southern California Law Review*, 93, p. 37.

KARPAN, A. (2018) *Troll Factories: Russia's Web Brigades*. Greenhaven Publishing LLC.

KAYE, D. (2019) *Speech Police: The Global Struggle to Govern the Internet*. Columbia Global Reports.

KEATS CITRON, D. (2019) Restricting Speech to Protect It', in: Brison, S.J. and Gelber, K. (eds.) *Free Speech in the Digital Age*. USA, Oxford University Press, pp. 122–136.

KELLER, D. (2020) 'Facebook Filters, Fundamental Rights, and the CJEU's Glawischnig-Piesczek Ruling', *GRUR International*, 69 (6), pp. 616–623.

KING, J., GOTIMER, K. (2020) How We Review Content. Available at: https://about.fb.com/news/2020/08/how-we-review-content/. (Accessed: 16 February 2021)

KLONICK, K. (2017) 'The New Governors: The People, Rules, and Processes Governing Online Speech', *Harvard Law Review*, 131 (6), pp. 1598–1670.

KLONICK, K. (2020) 'The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression', *Yale Law Journal*, 129 (8), pp. 2418–2499.

KOROŠEC, D., FILIPČIČ, K., ZDOLŠEK, S. (eds.) (2018) *Veliki znanstveni komentar posebnega dela Kazenskega zakonika (KZ-1)*. Uradni list Republike Slovenije.

KOVAČIČ, M. P. (2014) 'Etične Dileme Slovenskih Spletnih Novinarjev in Urednikov', *Javnost – The Public*, 21 (sup1), pp. 113–129.

KOVIC, M., RAUCHFLEISCH, A., SELE, M., CASPAR, C. (2018) 'Digital astroturfing in politics: Definition, typology, and countermeasures', *Studies in Communication Sciences*, 18 (1), pp. 69-85-69–85.

LI, Y., CHANG, M.-C., LYU, S. (2018) 'In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking', *arXiv:1806.02877 [cs]* [Online], Available at: http://arxiv.org/abs/1806.02877. (Accessed: 18 March 2021)

LINDIČ, L. (2017) 'Intenca strahu v sovražnem govoru: Premislek o vzrokih in posledicah sovražnega govora', *Časopis za kritiko znanosti*, (268), pp. 43–54.

LOBNIK, M. (2020) *Zagovornik načela enakosti – Redno letno poročilo za leto 2019*. Ljubljana, Advocate of the Principle of Equality.

LOTRIČ, S. (2021) *Poslanka SDS-a Alenka Jeraj bo po spodrsljaju izbrisala svoj anonimni profil na Twitterju* [Online]. Available at: https://www.rtvslo.si/slovenija/poslanka-sds-a-alenka-jeraj-bo-po-spodrsljaju-izbrisala-svoj-anonimni-profil-na-twitterju/570560. (Accessed: 6 April 2021)

MANCE, B. (2014) 'Odnosi Med Novinarji in Politiki na Twitterju', *Javnost – The Public*, 21 (sup1), pp. 23–40.

MASNICK, M. (2015) *Facebook Announces Its ContentID Attempt… Using Audible Magic* [Online]. Available at: https://www.techdirt.com/articles/20150827/16421932087/facebook-announces-contentid-attempt-using-audible-magic.shtml. (Accessed: 23 March 2021)

MILOSAVLJEVIĆ, M. (2016) *Je Trump res zmagal zaradi laži, ki so se pojavile na internetu, na Facebooku?* [Online]. Available at: http://www.mladina.si/177368/marko-milosavljevic/. (Accessed: 17 March 2021)

MILOSAVLJEVIĆ, N. (2010) Srbin ubio ajkulu ubicu u Šarm el Šeiku. Available at: https://www.njuz.net/srbin-ubio-ajkulu-ubicu-u-sarm-el-seiku/. (Accessed: 6 April 2021)

MORAIS CARVALHO, J., ARGA E LIMA, F., FARINHA, M. (2021) *Introduction to the Digital Services Act, Content Moderation and Consumer Protection* [Online]. Report no. ID 3852280. Rochester, NY, Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3852280. (Accessed: 12th July 2021)

Napoli, P. M. (2018) 'What If More Speech Is No Longer the Solution: First Amendment Theory Meets Fake News and the Filter Bubble', *Federal Communications Law Journal*, 70 (1), pp. 55–104.

Noor, R. (2016) 'Citizen Journalism vs. Mainstream Journalism: A Study on Challenges Posed by Amateurs', *Athens Journal of Mass Media and Communications*, 3, pp. 55–76.

Oblak Črnič, T. (2017) 'Fenomen ZLOvenija : v precepu med digitalnim aktivizmom in omrežnim individualizmom', *Časopis za kritiko znanosti*, 45 (268), pp. 181–198.

Oblak Črnič, T., Jontes, D. (2019) 'Digitalno nasilje v perspektivi medijskih študij: pomen medijske potrošnje in vsakdanjega življenja mladih', *Šolsko polje*, 30 (1–2), pp. 119–136.

Ombudsman RS (2021) *In Brussels, Ombudsman Svetina discusses the freedom of the media and the rule of law in Slovenia* [Online]. Available at: https://www.varuh-rs.si/en/news/news/in-brussels-ombudsman-svetina-discusses-the-freedom-of-the-media-and-the-rule-of-law-in-slovenia/. (Accessed: 9 April 2021)

Oversight Board (2021) *Case decision 2021-001-FB-FBR* [Online]. Available at: https://oversightboard.com/decision/FB-691QAMHJ/. (Accessed: 26 May 2021)

Pariser, E. (2012) *The Filter Bubble: What the Internet is Hiding from You*. Penguin Books.

Pavlin, M. (2021) *Cepivo ne povzroča neplodnosti. 'Za to ni popolnoma nobene osnove'* [Online]. Available at: https://www.24ur.com/novice/dejstva/cepivo-proti-covidu-19-ne-povzroca-neplodnosti-za-tem-ni-popolnoma-nobene-osnove.html. (Accessed: 2 April 2021)

Pelikan, E. (2015) 'Teorije zarote po slovensko: antisemitizem brez Judov', *Časopis za kritiko znanosti*, 43 (260), pp. 54–68.

Petrič, G., Rogelj, A., Petrovčič, A., Dremelj, P. (2015) 'Opolnomočenje V Spletnih Podpornih Skupinah Za Ljudi S Težavami V Duševnem Zdravju: Vloga Podpornega Komuniciranja in Motivov za Uporabo', *Teorija in praksa*, 52 (5), pp. 865–1004.

Plesničar, M. M., Šarf, P. (2020) ''This Web Page Should Not Exist': A Case Study of Online Shaming in Slovenia', in: Trottier, D., Huang, R.G. and Q. and Huang, Q. (eds.) *Introducing Vigilant Audiences*. Cambridge, Open Book Publishers, pp. 188–213.

Pohar, K. (2021) ''Clickbait' in nedovoljena uporaba podobe znane osebe', *Pravna praksa*, (11), pp. 27–28.

Prislan, K. (2013) 'Vpliv spletnih socialnih omrezij na dinamiko protestov', *Varstvoslovje*, 15 (1), pp. 19–28.

Pušnik, N. (2021) *Zloraba našega imena in ponarejanje anket javnega mnenja* [Online]. Available at: https://www.24ur.com/novice/slovenija/zloraba-nasega-imena-in-ponarejanje-anket-javnega-mnenja.html. (Accessed: 2 April 2021)

Reporters Without Borders (n.d.) *Slovenia : Threats, systematic smear campaigns and media concentration | Reporters without borders* [Online]. Available at: https://rsf.org/en/slovenia. (Accessed: 7 April 2021)

Reporters Without Borders (2021) *2021 World Press Freedom Index: Journalism, the vaccine against disinformation, blocked in more than 130 countries* [Online]. Available at: https://rsf.org/en/2021-world-press-freedom-index-journalism-vaccine-against-disinformation-blocked-more-130-countries. (Accessed: 21 April 2021)

Reynders, D. (2020) *Countering illegal hate speech online: Factsheet – 5 monitoring round of the Code of Conduct*.

Režek, M. (2010) 'Predgovor', in: Režek, M. ed. *Cenzurirano: Zgodovina cenzure na Slovenskem od 19. stoletja do danes,*. Ljubljana, Nova revija.

Ross, A. S., Rivers, D. J. (2018) Discursive Deflection: Accusation of 'Fake News' and the Spread of Mis- and Disinformation in the Tweets of President Trump. *Social Media + Society* [Post-print], 4 (2). Available at: https://doi.org/10.1177/2056305118776010. (Accessed: 21 April 2021)

Sánchez Nicolás, E. (2020) *EU warns tech giants 'time to go beyond self-reguation'* [Online]. Available at: https://euobserver.com/science/149404. (Accessed: 29 March 2021)

Šašek, N. (2020) *Porodničar, ki je objavil, da cepivo povzroča neplodnost, se je opravičil – 24ur.com* [Online]. Available at: https://www.24ur.com/novice/korona/porodnicar-ki-je-objavil-da-cepivo-povzroca-neplodnost-se-je-opravicil.html. (Accessed: 2 April 2021)

Savič, D. (2020) QAnon in predsednik slovenske vlade. Available at: https://www.dsavic.net/2020/07/22/qanon-in-predsednik-slovenske-vlade/. (Accessed: 2 April 2021)

Selak, Š., Kuhar, M. (2020) 'Socialno-psihološki vidik uporabe spletnega socialnega omrežja Facebook skozi prizmo dejavnikov selektivnega samopredstavljanja', *Teorija in Praksa*, 57 (2), pp. 622–645.

Selinšek, L. (2015) 'Veliko podatkovje v pravu in ekonomiji: veliki izzivi ali velike težave?', *LeXonomica*, 7 (2), pp. 161–188.

Silver, E. (2018) 'Hard Questions: Who Reviews Objectionable Content on Facebook — And Is the Company Doing Enough to Support Them? Available at: https://about.fb.com/news/2018/07/hard-questions-content-reviewers/. (Accessed: 16 February 2021)

Škerjanec, B. (2013) 'Virtualno vstajnistvo', *Časopis za kritiko znanosti*, (254), pp. 167–176.

Smokvina, T. K., Pavleska, T. (2019) 'Igra mačke z mišjo: medijska regulacija v Evropski uniji v času algoritmizacije komuniciranja', *Javnost – The Public*, 26 (sup1), pp. S82–S99.

Stajnko, J., Kičin, S., Tomažič, L. M. (2020) 'Kriminalitetnopolitično razumevanje 297. člena KZ-1 : sovražni govor in dva koncepta svobode', *Revija za kriminalistiko in kriminologijo*, 71 (1), pp. 31–41.

Stare, Š. (2020) *Statistika dela NČR za leto 2019*. Ljubljana, Novinarsko častno razsodišče SNS in DNS.

Statistical Office of the Republic of Slovenia (2020) *Uporaba interneta v gospodinjstvih in pri posameznikih, podrobni podatki, Slovenija, 2020* [Online]. Available at: https://www.stat.si/StatWeb/News/Index/9258. (Accessed: 16 February 2021)

Su, S. (2018) Update on Myanmar. Available at: https://about.fb.com/news/2018/08/update-on-myanmar/. (Accessed: 18 February 2021)

Teršek, A. (2018) *Svoboda izražanja, mediji in demokracija v postfaktični družbi: Filozofske, teoretične in praktične refleksije*. Ljubljana, GV Založba.

Teršek, A. (2019) 'Človekove pravice in temeljne svoboščine / 39. člen / Komentar: Opredelitev pravice in razmerje do drugih ustavnih pravic', in: Avbelj, M. ed. *Komentar ustave Republike Slovenije*. Nova Gorica, Nova univerza, Evropska pravna fakulteta.

Tushnet, R. (2019) 'Content Moderation in an Age of Extremes Essay', *Case Western Reserve Journal of Law, Technology and the Internet*, 10, pp. 1–19.

Vaccari, C., Chadwick, A. (2020) 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News', *Social Media + Society*, 6 (1).

Vaccaro, K., Sandvig, C., Karahalios, K. (2020) '"At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation', *Proceedings of the ACM on Human-Computer Interaction*, 4 (CSCW2), p. 167:1-167:22.

Valicon (2020) *Uporaba družbenih omrežij in storitev klepeta v Sloveniji 2018–2019* [Online]. Available at: https://www.valicon.net/sl/2020/01/uporaba-druzbenih-omrezij-in-storitev-klepeta-v-sloveniji-2018–2019/. (Accessed: 16 February 2021)

Vehovar, V., Povž, B., Fišer, D., Ljubešič, N., Šulc, A., Jontes, D. (2020) 'Družbeno Nesprejemljivi Diskurz Na Facebookovih Straneh Novičarskih Portalov', *Teorija in Praksa*, 57 (2), pp. 622–645.

Vidmajer, S. (2017) *Dom in svet: Lažne novice* [Online]. Available at: https://old.delo.si/mnenja/kolumne/lazne-novice.html. (Accessed: 17 March 2021)

Voh Boštic, A. (2021) *Astroturfing (1. del): lažni Twitter profili, ki napadajo kritike stranke SDS • Pod črto* [Online]. Available at: https://podcrto.si/astroturfing-1-del-lazni-twitter-profili-ki-napadajo-kritike-sds/. (Accessed: 2 April 2021)

Weingerl, P. (2020) 'Potrošnik kot uporabnik spletnih storitev', *Pravni Letopis*, 2020, pp. 201–218.

Whitten-Woodring, J., Kleinberg, M.S., Thawnghmung, A. and Thitsar, M.T. (2020) 'Poison If You Don't Know How to Use It: Facebook, Democracy, and Human Rights in Myanmar', *The International Journal of Press/Politics*, 25 (3), pp. 407–425.

Wiseman, J. (2020) New Administration, Old Agenda: Press Freedom Strained Again in Slovenia Under Veteran PM Janša. Available at: https://ipi.media/new-administration-old-agenda-press-freedom-strained-again-in-slovenia-under-veteran-pm-jansa/. (Accessed: 7 April 2021)

Yue, N. (2019) 'The 'Weaponization' of Facebook in Myanmar: A Case for Corporate Criminal Liability Notes', *Hastings Law Journal*, 71 (3), pp. 813–844.

Završnik, A. (2017) 'Zakaj obsodb sovražnega govora v Sloveniji ni?: od okov implementacije do rigidnosti razlage', *Zagovor javnosti : med svobodo izražanja in sovražnim govorom*, pp. 101–114.

Završnik, A. (2018a) 'Lažne novice in razpad realnosti', *Pravna praksa*, (15), p. 23.

Završnik, A. (2018b) 'Lažnivčeva dividenda in globinski ponaredki', *Pravna praksa*, (49–50), p. 25.

Završnik, A., Sedej, A. (2012) 'Spletno in mobilno nadlegovanje v Sloveniji', *Revija za kriminalistiko in kriminologijo*, 63 (4), pp. 263–280.

Završnik, A., Zrimšek, V. (2018) 'Sovražni govor po slovenski kaznovalni zakonodaji in sodni praksi : neustaven položaj', *Časopis za kritiko znanosti*, 45 (268), pp. 59–73.

Zobec, B. (2019) 'O 'prelomni' sodbi VS RS v zvezi s kaznivim dejanjem po 297. členu KZ-1', *Pravna praksa*, 38 (46), p. II–VIII.

Zupančič, A. (2020) 'Kratka razprava o teorijah zarote', *Filozofski vestnik*, 41 (3).

Zupančič, B.M. (2020) How George Soros' Open Society and Bill Gates' Microsoft fund the Council of Europe https://t.co/o5X0wpDQZc. Available at: https://twitter.com/bmz9453/status/1332101766192762882. (Accessed: 12th July 2021)

Legal sources

Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14*, 4 November 1950

European Union, *Charter of Fundamental Rights of the European Union*, 26 October 2012

European Union: Council of the European Union, *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law*, 28 November 2008

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA

Directive (EU) 2011/93 of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA

Directive (EC) 2000/31 of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (e-Commerce Directive)

Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services

Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC

Constitution of the Republic of Slovenia: Ustava Republike Slovenije (URS) Official Gazette 33/91-I, 42/97 – UZS68, 66/00 – UZ80, 24/03 – UZ3a, 47, 68, 69/04 – UZ14, 69/04 – UZ43, 69/04 – UZ50, 68/06 – UZ121,140,143, 47/13 – UZ148, 47/13 – UZ90,97,99 and 75/16 – UZ70a

Criminal Code: Kazenski zakonik (KZ-1) RS, Official Gazette 50/12, 6/16, 54/15, 38/16, 27/17, 23/20 and 91/20

Electronic Commerce Market Act Zakon o elektronskem poslovanju na trgu (ZEPT) Official Gazette 97/07, 64/16 – odl. US and 20/18 – OROZ631

Electronic Signature Act: Zakon o elektronskem poslovanju in elektronskem podpisu (ZEPEP) Official Gazette 98/04, 61/06 – ZEPT and 46/14

Enforcement and Security Act: Zakon o izvršbi in zavarovanju (ZIZ) Official Gazette 3/07, 93/07, 37/08 – ZST-1, 45/08 – ZArbit, 28/09, 51/10, 26/11, 17/13 – odl. US, 45/14 – odl. US, 53/14, 58/14 – odl. US, 54/15, 76/15 – odl. US, 11/18, 53/19 – odl. US, 66/19 – ZDavP-2M and 23/20 – SPZ-B

Mass Media Act: Zakon o medijih (ZMed) Official Gazette 35/01, 62/03, 113/03 – odl. US, 16/04 – odl. US, 123/04 – odl. US, 96/05 – ZRTVS-1, 60/06, 69/06 – ZOIPub, 36/08 – ZPOmK-1, 77/10 – ZSFCJA, 90/10 – odl. US, 87/11 – ZAvMS, 47/12, 47/15 – ZZSDT, 22/16, 39/16, 45/19 – odl. US, 67/19 – odl. US

Obligations Code: Obligacijski zakonik (OZ) Official Gazette 97/07, 64/16 – odl. US and 20/18 – OROZ631

Protection against Discrimination Act: Zakon o varstvu pred diskriminacijo (ZVarD) Official Gazette 33/16 in 21/18 – ZNOrg

Personal Data Protection Act: Zakon o varstvu osebnih podatkov (ZVOP-1) Official Gazette 86/04, 113/05 – ZInfP, 51/07 – ZUstS-A, 67/07, 177/20 – ZVOPOKD

Protection of Public Order Act: Zakon o varstvu javnega reda in miru (ZJRM-1) Official Gazette 70/06, 139/20

Radiotelevizija Slovenija Act: Zakon o Radioteleviziji Slovenija (ZRTVS-1) Offical Gazette 96/05, 109/05 – ZDavP-1B, 105/06 – odl. US, 26/09 – ZIPRS0809-B and 9/14
CJEU: Sotiris Papasavvas v O Fileleftheros Dimosia Etaireia Ltd and Others (C-291/13)
CJEU: Eva Glawischnig-Piesczek v Facebook Ireland Limited CJEU decision (C-18/18)
Constitutional Court of Republic of Slovenia: Up-614/15, 21.05.2018
Constitutional Court of Republic of Slovenia: Up-1019/12, 26. 3. 2015
Constitutional Court of Republic of Slovenia: Up-407/14, 14.12.2016
Constitutional Court of Republic of Slovenia: Up-407/14, 14. 12. 2016
Constitutional Court of Republic of Slovenia: Up-530/14, 2. 3. 2017
Constitutional Court of Republic of Slovenia: Up-417/16, 18.03.2021
Supreme Court of Republic of Slovenia: II Ips 75/2019, 06.02.2020
Supreme Court of Republic of Slovenia: I Ips 65803/2012, 04.07.2019
Supreme Court of Republic of Slovenia: XI Ips 40945/2018, 22.10.2018
The High Court of Ljubljana: II Cp 577/2019, 10.07.2019
High Court of Koper: Cpg 213/2017, 22.02.2018
High Court of Ljubljana: I Cp 2892/2017, 24.01.2018
High Court of Ljubljana: II Cp 701/2015, 20.05.2015
High Court of Ljubljana: II Kp 13079/2012, 30.10.2014
High Court of Ljubljana: II Cp 2066/2012, 30.01.2013
High Court of Maribor: I Cp 193/2012, 08.05.2012
High Court of Ljubljana: I Cp 1206/2015, 10.06.2015
High Court of Ljubljana: I Cp 2054/99, 14.12.1999
High Court of Ljubljana: II Cp 1666/2014, 05.11.2014
High Court of Ljubljana: II Cp 2634/2017, 17.11.2017
High Court of Ljubljana: II Kp 49761/2015, 24.08.2016
High Court of Ljubljana: VII Kp 56216/2017, 04.10.2018
Administrative Court of Republic of Slovenia: I U 1228/2011, 13.03.2012

Chapter VII

The Role of Social Media in Shaping Society

Aleš Rozehnal

1. Social Media and Freedom of Speech

The vast majority of citizens, and therefore voters, use social media as their primary source of information and news. The Internet and social media in particular are shaping our democratic dialog. Czech legal enviroment´s concept of freedom of speech is that everyone has the right to publish what they see as proper, and to forbid them from doing so would destroy freedom of speech. Freedom of speech is one of the basic features of a democratic society and one of the basic conditions for human development and personal fulfillment.[1]

This is not applicable only to information and opinions that are favorably received and rated as non-offensive or neutral, but also to those that attack or shock.[2] Therefore, this also applies to those that some online platforms describe as harmful. There is no reason to have freedom of speech that allows only speech that is not harmful. These are the needs of pluralism, tolerance, and free-thinking, without which there is no democratic society.

The view that absolute protection of freedom of speech is necessary does not lie in the naive notion that words cannot do harm, but in the belief that society benefits from the free flow and exchange of ideas, which outweighs the negatives caused by

1 Rozehnal, 2020.
2 Burto and, Jirák, 2001.

Aleš Rozehnal (2021) The Role of Social Media in Shaping Society. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 217–244. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_7

harmful ideas.[3] Any censorship is counterproductive because the truth will come to light only when conflicting ideas clash. Deleting posts or blocking users will not eliminate racism, xenophobia, indecency, or hatred, as this is ingrained in some people, and they also have the right to freedom of speech.

At first sight, the issue of freedom of speech appears trivial in the sense that it is either given or not. However, the situation is not that simple, because the freedom of speech has external limits, that is, those stipulated by regulations, and internal limits, which are immanent with the freedom, as it contains the liability for speech—not liability as a moral or philosophical category but liability as a category of law.

The external limits[4] of the freedom of speech in the Czech Republic include the provisions of the Criminal Code and the Rules of Criminal Procedure, the regulations providing for the protection of personal rights, the provisions regulating the content of advertisements, and a number of rules relating to electronic and digital media, such as the duty to refrain from jeopardizing the ethical, physical, and mental development of children and minors, and the duty to carry out impartial and balanced broadcasting.[5]

A huge public dispute arose when the Czech Parliament adopted a law restricting the freedom of expression in sensitive issues. Media named this act the *Muzzle Act*. The Muzzle Act was an amendment to the Rules of Criminal Procedure and the Personal Data Protection Act that prohibited the media from publishing any information that might serve to disclose the identity of an aggrieved person younger than 18 years of age, or the identity of a victim of murder, manslaughter, or certain other criminal acts causing grave harm to the victim's health, distribution of sexually transmitted diseases, certain criminal offenses aimed against woman's pregnancy, human trafficking, certain criminal offenses against human dignity in sexual life, and the criminal offenses of abandoning a child or person entrusted to one's care, battering a person entrusted to one's care, battering a person living in one's dwelling, kidnapping a child or a mentally handicapped person, and stalking. In addition, it was prohibited to publish images, video and audio records, or other information during court hearings or public sessions that would enable the disclosure of identity of the party aggrieved by the aforesaid criminal offenses.[6]

Another ban imposed upon the media, including social media, concerned the publication of information about any ordered or carried out wiretapping and recording of telecommunication operations, or the information retrieved there from, data on telecommunication operations or information obtained through surveillance of persons and items, provided that they allow the disclosure of identity of the person concerned, and provided that they were not used as evidence in court proceedings.

3 Ruling of the European Court of Human Rights in the case *Grinberg v. Russia*, Application No. 23472/03 dated 21 Oct. 2005.
4 Rozehnal, 2015.
5 Rozehnal, 2008b.
6 Rozehnal, 2020.

This information may be published only on the grounds of public interest if it prevails over the right to enjoy the protection of privacy of the person concerned.

A breach of this prohibition is subject to strict sanctions. The Criminal Code stipulates that a person who through negligence and without being authorized to do so publishes, advises, makes accessible, otherwise processes, or appropriates personal data of another person collected in connection with the execution of public authority (e.g., via wiretapping), and causes serious harm to the rights or legitimate interests of the person concerned, may be sentenced up to three years in prison or to a prohibition on undertaking professional activities. The qualified merits of this offense consist of the perpetration thereof via press, film, radio, television, and publicly accessible computer networks, such as social media, or via any other similarly effective manner.

Although the Muzzle Act has been reasonably modified, conformity with the Convention on the Protection of Human Rights and Fundamental Freedoms has been established only by reference to public interest. Worth mentioning is the provision of the Criminal Code that stipulates that whoever intentionally violates the protections of data, text, voice, audio, or video messages sent via an electronic communication network attributable to an identified subscriber or user who receives the message shall be sentenced up to two years in prison or punishment by disqualification.

Freedom of speech is a manifestation of will secured by all rights against slander, insult, abuse, etc. This means that the freedom is regulated by the law.[7] Social media cannot stand above the law, but they should have the right to publish what they want, even if they risk consecutive sanctions if any of them crosses the limits set forth. Social media users do not stand above the law.

Social media users are persons with the capacity to monitor suspicious behavior and instinctively gather information about things and topics that are not what they seem to be. Sometimes, they may be biased or in error, but this is a necessary consequence of the freedom of speech.

Observation of the limits of freedom of speech is supervised by the courts; however, it is not up to them to supplement the social media users' opinions and determine what techniques they should use. Judges must resist the temptation to become editors of social media posts. The court must not act as a censor.

Another pillar of freedom of speech is the fact that the government has no control over the media. The freedom of speech currently protects efforts to publish unpleasant information against the will of governments, multinational corporations, and public entities.

Multinational corporations often create more products and influence than many UN member states. More than half of the one hundred richest entities are corporations, not states. Some have breached the law or profited from such a breach.

The first limit of freedom of speech in social media thus rests on the definition of what is private and cannot be published. However, this issue may be a key subject in

7 Zelezny, 2001.

much litigation.[8] Certain matters in human life are considered private by nearly everyone, such as health conditions, marital issues, and crimes committed by children.

Another definition of privacy is that privacy is required wherever it may be reasonably assumed. The zone of privacy may be described locally as a crib, a school, a hospital, a toilet, a bedroom, and a grave. A democratic society must protect the privacy at least of those who do not commit any wrongful or immoral acts as part of their personal freedom and must offer them a choice of what aspects of their private life they wish to share with others.[9]

This freedom arises from the same source as the freedom of speech. Both the communist and Nazi regimes restricted the privacy of citizens in favor of the state through an apparatus of informants, agents, and censors. It is a task of democratic legal states to protect privacy. It is obvious, however, that not all states share this legal concept, even those we define as democratic.

Another limit of freedom of speech is the conflict of this right with the right to protect universal personal rights. Social media will always stand on the other side of the protection of personal rights, as many posts give a critical account of certain people, thus interfering with their personal rights. Therefore, it is important to find a balance between the two counter poles. In the event of a conflict between the fundamental political right to information and the distribution thereof, and the right to protect one's personal rights and privacy, that is, fundamental rights standing on an equal level, it will always be up to the independent courts to weigh the circumstances of each case and thoroughly consider whether one right was given unreasonable priority over the other.[10]

This is also stipulated under Article 4 Clause 4 of the Czech Charter of the Fundamental Rights and Freedoms, which imposes a duty on the bodies exercising the right to always consider the nature and meaning of the fundamental rights. The form, scope, and manner of interference with personal rights must always correspond with the purpose thereof, and the human dignity of the person concerned must always be treated with care; otherwise, it would constitute an unauthorized encroachment on the personality of an individual.

2. Democratic Censorship

In the Czech intellectual environment, the truth presented by social media is the truth that matters. If the social media says that an event or statement is true, it will be established as the truth, even if it is not. Therefore, the truth is what social media

8 Rozehnal, 2020.
9 Crone, 2002.
10 Judgment of the Constitutional Court of the CR, Case No. II. ÚS 2048/09 dated 2 Nov. 2009.

recognizes as the truth. The freedom of speech and the right to obtain information thus become imaginary because the only space for public discourse is in social media. The social media decides on the topics of the discourse as well as on the arguments and participants thereof. As opposed to autocratic censorship, democratic censorship is no longer based on omitting and deleting data but on the gathering, saturation, and surfeit of information. The information is now distorted by volume.[11]

Information is hidden or garbled because there is too much information, and the recipient does not even notice what is missing. One of the great differences between the world in which we have been living in the past decade and the world immediately preceding it is that information is no longer scarce.

In pre-modern times, whoever had information had power, which was understood as an instrument serving to control the circulation of information. Currently, the preponderant power is not in the hands of those who create the information but in the hands of those who distribute it, such as social media. The censorship in today's world looks different and has different intentions than in the past. It is based on more complex financial and commercial criteria, contrary to authoritative censorship. The flood of information masks the lack of relevant information and obscures the fact that the images are often false and actually conceal reality.

This came after enchantment with the media in the 1970s and the 1980s when the media, being the "fourth power," were presented as a prospective refuge from the misuse of the other three powers (executive, legislative, and judicial) and as a civic guarantee of true democratic control. Journalism was adored as independent, fair, honest, and strict. It defied general decline and seemed to be an authentic knight of the truth and a loyal ally of the helpless citizen.[12]

Thus, the media were defined as the fourth power. For us to speak of the "fourth power," the three other powers would have to exist, along with the hierarchy according to which Montesquieu classified them. At present, the state is becoming somewhat emptied, as it leaves some of its functions to influential and economic entities. In fact, the premiere power today is the entrepreneurial segment. The second power (which seems to be strongly associated with the first one) is certainly the power of social media as a tool of influence. Political power stands third. Therefore, the state is no longer the greatest power in society, which is manifested by the fact that corporations have a higher rating than the state itself.

Until now, the news relationship was based on a triangle of three poles: an event, a journalist, and a citizen. The event was passed on to the journalist for verification, re-filtering, and analysis, and then passed on to the citizen. Now, the triangle becomes a line segment with an event on one end and a citizen on the other.

The journalists' functions have disappeared. There is no longer a filter or sifter in the middle. Social media tries to bring citizens into direct contact with the event through a camera or written news coverage. These principles of the news function

11 Rozehnal, 2015.
12 McLuhan, 2003.

make it very difficult to realize the fact that the information means freedom, which means democracy.

Another auto-censorship element is the unilateral orientation of the social media toward negative news. This trend is somewhat understandable, as it is closely tied to the critical nature of the media, which is imminent. On the other hand, if social media creates a negative image of the world, this image becomes a tool of indoctrination and manipulation of the public because of its natural desire to find a Messiah who would change the negative world for the better.

Therefore, a confrontational style is the only form through which social media operates. The principle of discussion and dialog is thus entirely abandoned, as social media believes that the escalation of a dispute is much more attractive. This tendency establishes a certain model of communication that politicians consider necessary.

Another source of problems is the continuous acceleration of the circulation of information and the great demand for more information. That is, the basic criterion of the news should be veracity, impartiality, and balance, as well as the speed of the news transmission from the social media to the recipient.[13] The speed then places a great demand on social media, which has no time left to verify the information for truthfulness, completeness, impartiality, and balance. News is a commodity that spoils fast, and any delay in the publication thereof may reduce its value and public interest. Therefore, a certain error in the presentation of the news caused by speed must be permitted, as well as exaggeration or provocation.

It is necessary to respect certain specific features of the social media designated for the distribution of information to the broad public (as opposed to traditional media or professional publications, for example), which in certain cases must simplify, namely with respect to the scope of individual social media users' interest.

It cannot be stated without explanation that each simplification (or distortion) must necessarily lead to interference with the personal rights of the persons concerned. We could hardly insist on the absolute accuracy of facts and thus impose impossible demands on social media. It is always important that the overall impression is that the information corresponds with the truth.[14] To review whether the right to human dignity and honor was infringed, the information published on social media concerned should be examined based on the following aspects:

(a) The seriousness of the assault, with respect to the fact that the greater the impact of the false assault of dignity and honor, the more misinformed the public.
(b) The nature of the information and scope of public interest.
(c) The source of the information, namely with respect to whether the sources do have direct knowledge of the event and whether a mere personal animosity or an attempt to gain profit is involved (such information may be true, but if the provision thereof is based on personal animosity, the social media

13 Rozehnal, 2020.
14 Pember, 2001.

user is more obliged to verify the information. Social media users may also sometimes trust a source who wishes to remain anonymous for fear of harm, that is, if the source is based in a country ruled by a repressive regime);

(d) The status of information, namely with respect to whether it is provided by a bearer of public power or authority.
(e) The steps taken to verify the information
(f) The urgency of a given matter.
(g) The overall impression made by the post, namely with respect to the fact that the information presented in the post should not be overstated or sensationalized, that speculation, and rumors should not be presented as facts, and guilt should not be presumed in advance.
(h) The circumstances under which the news is published, including the timing thereof.

Social media users' conduct must also be judged in the context of their post-publication acts, that is, whether they remove an error, explain an issue, or offer an excuse. It should also be kept in mind that opinions on virtue and dignity change. Is it defaming these days to say that someone is gay? Is it offensive to say that somebody is ugly? An offense is judged according to an average reader or a viewer who can tell and understand irony.

3. Conception of the Presumption of Innocence in Social Media

The greatest burden in terms of the protection of personality lies in the coverage of criminal cases. A criminal case is usually an issue of legitimate public interest; on the other hand, it is necessary to be mindful of the personal rights of individuals. Coverage should always be consistent in distinguishing the individual stages of a criminal procedure and pre-hearing stages, and should always take into account the presumption of innocence.

If somebody is sentenced for a criminal offense and the judgment is not final and conclusive, the coverage should inform readers of this fact, pointing to the fact that the judgment may still be appealed. In such a case, the social media should also report the result of the appeal to the same extent as it reported the preceding stages of the criminal procedure.

Therefore, an individual who has not been sentenced by a final and conclusive judgment should always be referred to as a suspect, defendant, or accused person, not as a perpetrator, murderer, thief, swindler, or rapist.

The presumption of innocence is breached when a court ruling, or another decision of a public body, condemns an individual as guilty without culpability being

proved by legitimate means. The presumption of innocence may also be breached by a statement from which it can be deduced that a court or another public body considers an individual concerned guilty.[15]

What is absolutely unacceptable are speculations in the social media about the culpability of an individual who has not been sentenced by a final and conclusive judgment, as such a judgment falls under the exclusive competence of the court and any speculations of the media concerning culpability prior to the conclusion of the proceedings are not only unlawful or unconstitutional but also unethical and unprofessional.

However, respect for the principle of the presumption of innocence is much more difficult for media than for the justice system, as they have a much wider scope of expression than the courts that only find guilt or innocence.

In contrast, media works with images and emotions and can therefore evoke the idea of guilt or innocence even with the use of relatively correct verbal means. A criminal charge means a great interference in everyone's personal integrity, professional, family, and social life, but often also in their health.[16]

We interpret the principle of the presumption of innocence as meaning that everyone should be deemed innocent until convicted of a criminal offense. For media, the interpretation of the principle should be somewhat shifted so that everyone should be seen as to have been eventually acquitted.

This holds not only in terms of criminal law, but also in terms of morality, society, and general reputation. The only harm that the person in question must tolerate is that which is imposed upon him/her after the crime has been committed.

If a criminal case is of interest to the media, the person concerned is often convicted in the eyes of the public before it is established that the crime has actually been committed. The person is *de facto* condemned regardless of whether he/she has committed the crime or not.

The question arises whether it is the duty of a civilized society to deprive a person of his/her position, employment, and ability to discharge offices and make it more difficult for him/her to succeed in society or endanger his/her family life as soon as charges are brought against him/her.

The ostracism associated with this is a punishment on a moral as well as a legal level, despite the fact that so far no punishment has been imposed by law. It is much more difficult then for defendants to find employment, establish themselves in society, and establish interpersonal relationships.

The principles of the rule of law should not only be applied in court proceedings, but should permeate society as a whole and all social relations, including the activities of media. Values such as correctness, trustworthiness, respect for human personality, and the presumption of innocence should not be confined solely to a courtroom. In this respect, however, our media are failing.

15 Rozehnal, 2020.
16 Knap et al., 2004.

There are opinions that social media should be banned from reporting criminal cases at the pre-trial stage. However, criminal cases are mostly a matter of legitimate public interest. It is thus necessary to find a kind of *modus operandi* to protect the rights of the persons concerned.

We are witnessing a growing appetite by the repressive forces of a state for power. This is not surprising, as it is a natural feature. Whenever repressive forces have the opportunity to seize more power, they will do so. This is also the reason for the existence of the division of power, the system of checks and balances, and control of power.

There should also be control over social media power. However, social media in the Czech Republic seems to have completely lost its democratic instincts and is essentially becoming an extended hand of repressive forces.

The situation is in many ways reminiscent of the period of McCarthyism in the 1950s in the United States, when there were certainly Soviet spies in the United States and some media outlets made a good effort to draw attention to their work. However, the result of this activity was shameful, and the country subsequently came to the brink of a Cold Social War.

There are several non-verbal media elements that run counter to the principle of the presumption of innocence, but these are widely used by our media. This is even though the media mostly adhere to the principle of the presumption of innocence, and some, especially public ones, even have it embedded in their code of ethics, in practice it violates this principle.

Here, I believe that if self-regulation fails, regulations should be imposed. For example, it is incompatible with the presumption of innocence to take pictures of detainees with handcuffs, because such a presentation of people has the same goal as centuries ago—to show a person enslaved, degraded, broken, defeated. However, this practice does not belong to a civilized society.[17]

The presumption of innocence is violated even if social media presents the opinion of a certain state authority on the guilt of the person concerned, although his/her guilt has not yet been legally established. This also applies if the opinion of the police or the public prosecutor's office is presented in this way, regardless of whether their subjective beliefs are different.

Another phenomenon that undermines the presumption of innocence is information asymmetry. It is based on the media's dependence on a single source, which is inherently biased. This source is usually the police or the prosecutor's office, which submits its version of the story. The defendant, on the other hand, is often prohibited by law enforcement authorities from reporting on the case, which further weakens the possibility of making at least a lay judgment on what actually happened.

We are certainly aware of the dangers of spectacular social media pseudo-judgments. In them, the nation acts as a jury, before which, however, the accused or

17 Rozehnal, 2020.

charged person does not have the opportunity to defend himself/herself or bring evidence regarding what he/she is accused of.

The principle of the presumption of innocence is not easily copied into the work of social media, as there could be virtually no investigative shows. However, this criminal law principle must be applied, because otherwise the repressive elements of the state will end up deciding which people are *de facto* enjoying full legal protection and who is to receive only semi-legal protection; media will multiply this interpretation until, as a result, everyone will be lawless.

It is also characteristic of our society that there is a kind of deification of criminal law, as if perhaps all the problems of society could be solved by criminal law. However, if society cannot cope with its day-to-day operations other than with the help of the police and criminal law, it is a sick society.

If social media supports the extension of this criminalization, it leads to social tensions, the suppression of human rights and freedoms, and the stigmatization of a wide range of people for whom the criminalization of their actions can be socially and humanly destructive.

One of the functions of the state in ensuring justice should also be to protect the individual from the mood of the masses. The masses are often not interested in the administration of justice but in the satisfaction of its baser instincts, which, however, are very far from the principle of justice.

If a viewer sees on television or reads in the newspaper about a defendant or an accused person who does not have the opportunity to defend himself/herself, he or she will convict the accused person, regardless of how the case will actually turn out. The person concerned is socially condemned, whether he/she is guilty or not. The concept of the presumption of innocence is replaced by the concept of inquisitorial instruments.

Social media plays a key role in shaping attitudes, opinions, beliefs, and values in society. However, if the public does not learn to understand the true nature of the presumption of innocence, then the criminal justice system alone cannot fulfill its purpose and cannot protect society from the real perpetrators of crime.

It is quite natural that the mood of society, including social media, is always directed against the accused and that the public and media have a greater tendency to trust the police and public prosecutors' offices than accused or charged persons.

It is dangerous if social media stirs up these tendencies even more and takes acquittal as a failure of the justice system. This is again a question of information asymmetry. We are always influenced by one side of the story by one source of information (law enforcement authorities), and the accused or charged person is usually forbidden from providing information or comments on the case.

If there is an acquittal, there is a huge difference between what society was informed about at the beginning and how the case itself turned out. This causes frustration within society and, at the same time, the feeling that justice is weak in the most serious cases of crime.

If society is convinced that there are criminals among us who are caught by the police but not punished by the judiciary because they cannot do so, are corrupt, or

succumb to criminal-minded lawyers who are involved in criminal proceedings, then they will lose faith in the criminal justice system itself.

At the same time, social media play a very important role in maintaining social cohesion. However, this cannot exist without a functioning justice system or its positive perception. The justice system must have the trust of society; otherwise, it cannot function properly and becomes illegitimate to some extent.

All undemocratic regimes, states, and governments have always used criminal law as a tool to enforce their will, even if the rule of law has been maintained. Democratic states are also taking these steps if they want to punish their (albeit often supposed) enemies.

Therefore, sufficient guarantees and means of protection for individuals in criminal proceedings must be provided. However, legal guarantees are only a necessary minimum. The superstructure is control by the public through social media over how the principles of criminal law are implemented. Public scrutiny is a prerequisite for fair justice and is exerted through the media. Therefore, for social media to perform control, these principles must be adopted.

So, in my view, although it is not very popular, even enforcing respect for the principle of the presumption of innocence in the social media should be a matter of legal regulation.

Before the final ruling is passed, social media users should refrain from any comments or deliberations because they could influence the decision-making of the court. It should also be prohibited for the media to challenge or question the final verdict of not guilty.

However, if a certain rule of law or a method of its application results in a gross discrepancy with the general understanding of justice, the critical evaluation of such a fact or such an application is acceptable.

4. Hate Speech

Most democrats are supporters of the widest possible freedom of speech because democracy cannot exist without freedom of speech. However, supporters of absolute freedom of speech are, paradoxically, most often recruited from racists or propagators of pornography, whom we can hardly call democrats.[18]

Many rules apply to the right to express facts and opinions in public debate, so their expression will never be completely free. Certain information, in fact, can cause irreparable damage, such as disclosing military classified information, influencing courts of law, or inciting racial hatred.

18 Rozehnal, 2020.

Thought must always be free, but communication of one's own ideas and interpreting them to someone else may be subject to restrictions. However, such restrictions must always be law-based, and every restriction must be clear, specific, and predictable.

Therefore, the formulation of laws has a precondition that the laws be adequate or comprehensible so that citizens may follow them. Although society urgently needs legal restrictions, they must always be adequate and satisfactorily justified by public authorities.

Freedom of speech, which is essential for a free state, means that public authorities do not put any preliminary obstacles in the media. However, this does not mean that they cannot apply any relevant restrictions or sanctions.

Everyone has the right to publish whatever they consider appropriate; this is the essence of freedom of speech. However, if they publish anything against the law, they must accept responsibility for their actions. Freedom of speech is one of the basic pillars of a democratic society and one of the main conditions for self-development and self-fulfillment. This applies not only to information and opinions that are well-received and judged as non-aggressive or neutral, but also to those that are aggressive, shocking, or irritating.[19]

The view that the greatest possible protection of freedom of speech is necessary does not lie in the vague idea that words cannot cause harm, but in the common belief that the free flow and exchange of ideas is beneficial to society, outweighing negative issues caused by "harmful" ideas.

Therefore, it is not appropriate to consider whether a strident activist should be allowed to protest abortions in front of abortion clinics. It is also wrong if a court of law punishes someone who has simply drawn tentacles on a politician's poster because it is an expression of a political opinion that is not harmful to society, even though it is a bit childish.

Freedom of speech and expression is the free market of ideas, where false, criminal, and harmful doctrines will be overcome by true statements and right opinions. This freedom of the free market of ideas and the defeat of false ideas is beneficial to society as a whole.[20]

However, the issue is not so simply resolved because freedom of speech has its limits, and not only the outer limits of legal regulation, but also inner limits immanent to this freedom, because freedom of speech also includes responsibility for the speech, which does not mean moral or philosophical responsibility but legal responsibility.

Internet use has brought a new dimension to the expression of freedom of speech. Easy dissemination of information in cyberspace is such a change in the amount of information as to constitute a change in quality, and thus a change in the understanding of freedom of speech. Anyone can speak to a large number of people

19 Rozehnal, 2011.
20 Drgonec, 2013.

with no physical or mental effort while receiving immediate responses. Moreover, this is fully or partially anonymous and, in contrast to the publication of articles in standard media that are highly elitist, extremely plebeian.

At first sight, free and universal access to the Internet has enabled as many people as possible to place their own ideas before the social consciousness; for the first time in history, everyone has the same opportunity to both accept and disseminate their points of view, and thus to participate in free civil society life. Thus, the Internet is a highly democratizing media environment, which is supposed to strengthen human rights and civil liberties.

Instead, we witness a mass of hate speech on the Internet, which is mainly posted in Internet discussions related to published issues, especially on news and journalistic servers. This hate speech is usually so severe that it interferes with the personal rights of other people in discussions or people discussed in the main issue, or it is so rude and vulgar that it violates the basic rules for civil coexistence. Its creators are often racist and xenophobic, proclaiming intolerance and contempt for democratic systems and other people's rights.[21]

The question is whether the organizer of the discussion, which is mostly an information service provider, should filter, delete, or otherwise interfere with such hate speech. It is evident that such interference is a limitation to the freedom of expression.[22] As mentioned above, a restriction of freedom of speech is possible only if it is permitted by law, if necessary, in a democratic society and if done in order to protect the values of a democratic society, that is, to protect the rights and freedoms of other people, eventually in order to maintain morale.[23]

Therefore, the question is whether we should fight hate speech on social media on a basis different from the legal basis, particularly on an educational basis. It is possible that the criminalization of a certain expression only hides the true nature of a problem to which society cannot respond. At the same time, such a restriction on hate speech initiators may increase their radicalization and drive them into a ghetto, which may result in social riots. Additionally, restrictions on freedom of speech lead to the distortion of democracy, which should not resolve conflicts by violence but through debate and persuasion.

As for the above-mentioned considerations, it is essential to the form of democracy we prefer. It is obvious that most of those initiating hate speech on the Internet are opponents of democracy. In the past, it often happened that opponents of democracy were able to use democratic means to gain power and subsequently destroy it, which democracy was not able to prevent in any way.

On the other hand, there is a philosophical and legal model of democracy fighting back, which entails the protection of democracy from its opponents, even at the cost of suppressing the basic principles that it is based on. Such defense must be as strong

21 Rozehnal, 2015.
22 Drgonec, 2008.
23 Rozehnal, 2020.

as its enemies, who are ready to destroy democracy. Our Constitutional Court agreed with this concept, stating that its "*legal use is legitimate considering the historical experience with Nazi and communist totalitarianism not only in our state, but also in the European context. If opponents of democracy and the values that democracy is based on are prepared to attack it, then the democratic system must also be prepared to defend itself against such attacks, including, if necessary, restrictions of fundamental rights*." The European Court of Human Rights has also accepted this concept as the principle of European democracies.[24]

Therefore, if organizers of discussions on the Internet want to contribute to the development of a democratic society and thus fulfill one of the main tasks of the media, they must remove hate speech. Irrespective of the somewhat dogmatic interpretations of the Information Society Services Act, some people believe that they should not do so.[25]

The fight against hate speech should take place at a level other than the repressive.

5. Fake News

The first failing of the law regarding the media environment is that the law did not respond to the problem of fake news. The problem of fake news is perhaps even greater than we realize, essentially constituting an information war.

The most read fake news articles on Facebook were read and shared more than any articles from mainstream media. At the same time, Facebook is the most common source of information about the government and political situation among millennials.

The possibilities of the Internet have brought a new dimension to the expression of freedom of speech. The ease of disseminating information in cyberspace represents such a change in the quantity of information that, in its consequences, it has led to a qualitative change, and thus to a change in the understanding of freedom of speech.

At first glance, free and universal access to the Internet has made it possible to promote as many people's opinions as possible to social consciousness; for the first time in history, everyone has been given the same opportunity not only to accept but also to disseminate their views and thus actively engage freely in civic society's life. The Internet is therefore a highly democratizing media environment, which should increase the level of human rights and civil liberties.

However, reality is different. Instead of democratization, the Internet has made it easier to spread hate speech and fake news. The law is helpless in the face of

24 Radbruch, 1999.
25 Rozehnal, 2008a.

fake news, and that helplessness begins with the fact that we cannot define what fake news really is. Fake news is different from biased, unbalanced, or inaccurate news.[26]

These include cases where the title does not correspond to the text of the article, cases where true content is disseminated with false contextual information, and cases where false or misleading content is disseminated. Fake news has so many different meanings that the term is becoming worthless.

It is very difficult to determine the line between fake news and an alternative point of view. The argument that most people can figure out fake news when they read it can disarm us in face of a threat to freedom of speech.

The danger of fake news is that it devalues and delegitimizes the information and views of real authorities and the concept of objective data, and weakens society's ability to make rational, fact-based decisions that contribute to social chaos. "Fake news" weakens trust in social institutions and, moreover, it is almost irrefutable by true information.

The spread of fake news has consequences. Democracy is based on people making decisions and choices based on information. People do not have to be experts, but they must have basic knowledge of the world in which they live. If their knowledge is twisted, they make bad decisions.

The exchange of information is democratized thanks to social networking platforms and digital content production technologies, such that everyone can create seemingly credible information waste that is difficult to distinguish it from quality information. Demand for "fake news" is a natural by-product of a faster news cycle and greater demand for short-format content.

Some jurisdictions have taken the path of labeling sources of fake news. Distinguishing between traditional and respected journalism and "fake news" sources is not very effective. Creating blacklists of websites can be even more dangerous than blacklisting scientific predatory journals, as it could lead to only government-approved sources of information being on the right list.

People now trust more personally communicated information than that provided by authoritative sources, often referred to as the guardians of the gateway through which information was released to the public in the past. Traditional gate keepers are less effective and visible. In addition, "fake news" is often presented as traditional journalism, the importance of which diminishes.

The current guardians of the gate are perceived not as providers of public service, but rather as entrepreneurs with information whose main goal is profit. Fake news is a symptom of deeper structural problems in the media environment. People believe more in what others in their circles promote or share. If a crowd starts running, one instinctively starts running as well. Historically, humans to protect themselves against predators in this way, but in today's digital age, contrariwise, it makes humans vulnerable.

26 Rozehnal, 2015.

Blocking fake news sources is also not an answer. In addition, blocking creates polarization. One can also be of the opinion that the law provides protection against fake news through the right to protection of personality, because fake news is mostly offensive and defamatory.

Suing for personality protection is not very effective. It is expensive and exhausting, with very uncertain results. It is often not even clear whom to sue. Actions against platforms that disseminate fake news are ineffective because of the legislation contained in the Act on Certain Information Society Services.

Earlier efforts to influence the public with one-to-many content-based technologies have been replaced by social networking technologies that allow propaganda to be aimed at targets who are more likely to adopt and further disseminate the propaganda content.

Fake news is sometimes disseminated without malicious intent, forwarded via social networks without the user checking the content. Sometimes they are taken over by journalists who are under pressure from social networks that disseminate information in real time.

Everyone plays a key role in the fake news system. Whenever a user receives or shares information without verifying it, this contributes to contaminating the media space. This space is now so polluted that everyone has a responsibility for what they receive and spread through cyberspace.

The law should focus on methods of distributing information, rather than content. However, such regulation will still interfere with freedom of speech. Restriction of freedom of speech is possible only if it is permitted by law, if it is necessary in a democratic society, and if it is done in the interest of protecting the values of a democratic society, that is, in the interest of protecting the rights and freedoms of others or maintaining morale.

At the same time, protecting a democratic system of law against those who would threaten it cannot be allowed to narrow the limits of freedom of expression too much. No legal norm will eliminate racism, xenophobia, indecency, or hatred, because they are rooted in some people who also have the right to freedom of speech. If the courts make rules regarding what speech is still acceptable and determine what is already in conflict with law, they favor the majority opinion and ostracize the opinion of minorities.

6. The Concentration of Social Media Ownership

Another problem to which the law has not provided a satisfactory answer is the concentration of media and political power.

The Czech Constitution is based on the classical division of power into legislative, judicial, and executive branches, among which it seeks to create a system of checks

and balances. However, this tripartite power is supplemented by the Constitution, as well as by other laws regarding bodies with other powers that do not fall under the standard triad. These other powers include, for example, the power exercised by the Czech National Bank, the Supreme Audit Office, and the Public Defender of Rights.

The Constitution also sets out the legal framework for the functioning of the so-called fourth power of the media. Social media is not only one of the other products on the market, but it also has constitutional value. The Czech Republic is defined as a democratic state. Democracy can survive long social and economic difficulties, but it will not survive without free and independent media.

Social media plays a key role in shaping the attitudes, opinions, beliefs, and values of society, playing a political role and exerting political effects. Social media influences politics, political processes, and the electorate. Social media content has a significant impact on people who consume media content.

Using social media is a practical exercise of the freedom of speech of every citizen. Citizens have the right to comment on matters of public interest through social media. Publicity is the soul of democracy, guaranteeing that democracy can exist. Freedom of speech is one of the basic features of a democratic society and one of the basic conditions for human development and personal fulfillment.

Free and diversified ideas in social media are a vital component of a healthy democracy, because only the media are able to convey the opinions of citizens to those who will rule on their behalf, that is, the political elites they elect. Media ownership and control have implications for the nature of public debate, people's attitudes toward social issues, and social conflicts. The problem of who controls social media is, therefore, a fundamental problem in a democratic society.

Social media ownership can affect its content and, therefore, the political development of the country. This development is then conditioned by the existing political and regulatory institutions in the country. If there are strong democratic institutions in the country, the tendency to concentrate media power is obviated by these institutions, often non-governmental ones. If an oligopolistic media environment is created, there is less chance that the media will approach social problems independently. In extreme cases, an oligopoly can prevent the control or criticism of state power by independent media.[27]

The constitutional value of social media goes far beyond the financial interests of individuals and societies. Social media is not just one of the other products in the market. Democracy can exist without 15 different types of margarines; it can survive long social and economic difficulties; but it will not survive without a free and independent press. The press cannot be replaced by anything. If the media were to be controlled by individuals involved in political power, especially the executive branch, this would have a devastating effect on media pluralism. The combination of political and media power is at odds with our perception of the democratic functioning of the media.

27 Rozehnal, 2015.

At the same time, the importance of Internet news servers is equal to that of television news, and for citizens, the Internet is becoming one of the main sources of information. This trend is likely to intensify in the future, as the ever-growing digital world affects almost every aspect of our lives. A growing number of digital communication platforms, such as social media, blogs, and websites, is quickly and easily accessible. We are starting to depend on them not only in business, education, and personal life, but also in other areas.

Local media now compete globally, and domestic consumers have access to international content and services. Convergence is changing established investment patterns, competition, and the structure of the media market. Media distribution is no longer tied to a specific network, thus increasing the availability of these services and products.

This changes the relationship between the content creator and the content distributor. Content creators have more options to distribute their content and can reach readers or viewers through a variety of platforms. Distribution thus ceases to be a barrier for content creators. Traditional distributors have less influence on the habits of media consumers, as they can choose between different distribution methods. The way content is created is also changing. Audio-visual content is no longer produced by large professional media companies. Platforms such as YouTube or Stream are gaining popularity, which means that everyone can create and share content very easily and quickly.

Another change is that jurisdictional boundaries also play an ever smaller role, as digital data is potentially available anywhere. For a long time, there was the idea that the Internet is an extra-lege environment and thus a kind of shield against legal regulation. Cyberspace is a social space, and the same rules apply there as in any other social space, including the rule of law. However, actions in cyberspace are not different from the actions we know from other media. There is no reason why cyberspace should be immune to classical regulations.[28]

Before trying to find an answer to the question of whether the concentration of social media power, or more precisely, the concentration of media ownership is dangerous, it is necessary to answer the question of what the role of the social media actually is.

Whenever we discuss democracy, we pay attention to the state's institutions, how elections are organized, the structure of state authorities, and the form of the government. However, democracy does not simply mean what happens in a state. It includes a wide range of issues concentrated in the public sphere of life. It is the sphere of social life outside the state where people come together to discuss various issues, such as political views. The key question of such a social life is the question of information.

Social media produces content, which is called a public good in economics terminology. A public good is considered a non-competitive good in economics, the use

28 Rozehnal, 2015.

of which by a person does not exclude other people from using it. Public goods are offered collectively, and they are mostly funded through taxation. Financially, the media have some specific characteristics that are different from those of other public goods.[29]

Social media plays a significant role in forming attitudes, opinions, beliefs, and values in our society, as well as having a political role and exerting a political influence. Social media influences politics, political processes, and the electorate. Social media content has a significant impact on the people who consume media content. In economics terminology, we call this influence an externality, which can be positive or negative.

Perhaps the social media would not be successful in telling people what they should think, but they are in fact successful in doing so. The agenda-setting of social media means that social media allocates different levels of public attention to news topics. If social media ignores a topic, it is considered less important by the public, and if social media pays more attention to a topic, it is considered more important.

Social media allows citizens to come together and compare their political views with those of others. The media are also mediators between citizens' political and private spheres. To the extent that citizens are well-informed, they can judge and fit into different parts of the political spectrum. Social media also plays an important role in maintaining social cohesion as a mirror of society as a whole; they not only affect what we think about things overall but also what topic we think about.

Communication is the main source of human interaction. While small groups of people can communicate face-to-face with each other, society itself depends on free and independent media to ensure the exercise of the freedom of expression and information. Media freedom includes the right of the public to a media freedom system that offers balanced, complex, and diverse information. Such a system of media freedom is a basic necessity for an effective democratic system. Without free media, there would be neither free and unbiased information nor a public debate on social life issues.[30]

Society is becoming increasingly dependent on information. There is therefore a special responsibility and power of social media. The free market of ideas serves the public interest by maximizing the chances that lies, and misinterpretations will be revealed and that citizens will hear both sides of arguments and form their own well-informed opinions. If there is no free market of ideas and information, the public will not receive the necessary information important for their self-government.

However, this view has several limitations. While it is very unlikely that the political views of an electric kettle manufacturer will influence the political views of

29 Rozehnal, 2020.
30 Rozehnal, 2008a.

his or her consumers, it is highly likely that political views of social media operators will influence the nature of the information that the social media produces. Typically, social media operators influence neglected spheres, for example, by choosing banned posts.

Another problem is that the social media market is highly competitive, and barriers to entry exist. The Internet is full of new blogs and sites, some of which are very popular. Sometimes, the web plays an important role in discovering information of current importance. The Internet may be a counterweight to big news organizations. It is extremely expensive to gather information, online news portals are connected to large media companies, and those that do not lack the resources to fund news.

A very important role of social media is its watchdog role in democracy. Through this role, they are irreplaceable and represent the public in the supervision of state power. The media business is not like other businesses because of the nature of the media's product, and journalism is unlike other areas of business.

Media communication, ownership, and control affect the nature of public debate and people's attitudes toward social issues and conflicts. How the media are controlled is therefore a crucial problem for a democratic society. Therefore, social media control is a key problem in media and democracy.

Although everyone agrees that free media are utterly necessary to a free society, there are relatively many ambiguities in what the word "free" means. Generally, it means that the media should be free of government regulation, which means free market media. However, free media must be free not only from state power but also from ownership power.

One of the main problems with the media in relation to their ownership is whether they will serve the ideas of democracy or their owners' interests. Social media production can be considered a business, but successful business metrics are different from metrics of successful democracy. The conflict of interest is inherent in private media ownership, which can lead to a situation where the social media will stop playing the role of a democracy's watchdog or become a watchdog that does not bark.[31]

Thus, there exists an essential conflict between the economic nature of social media as a business and the social and political roles that social media is supposed to play. Of course, social media operators have their own, often different interests, which take the form of political or economic motivation. This motivation determines the content of social media. If the motivation is political, the social media support a specific political agenda; if it is economic, then the main goal is to raise funds. Thus, there exists a latent conflict between corporate and private media ownership and the principles of democracy.

Such a permanent conflict between the right of a social media operator to act opportunistically in response to changing market conditions and freedom of social media users is premised on the concept of the human right to freedom of expression.

31 Rozehnal, 2015.

This contradiction is a basic dilemma in the social media business. Various attempts have been made to solve this dilemma, but there has been little success. In the end, the right of a social media operator to determine the line of social media has always prevailed through the right of ownership. However, if social media serve commercial or political interests, they often forget their roles of monitoring and controlling the elected representatives of society.

Social media ownership can affect the content and, thus, political development in a country. Development is conditioned by the existing political and regulatory institutions in the country. If there are strong democratic institutions in a country, then tendencies to concentrate media power are controlled by democratic institutions.

Democracy is an endless process of cognition and education. The risk of censorship is still raised in Western democratic countries without any significant political changes, such as, for example, in Czechoslovakia in February 1948.

The concentration of social media ownership has long since been considered a major threat to media pluralism and diversity. The concentration of social media power is unequal and therefore undemocratic, uncontrolled, and potentially irresponsible.

The Czech constitutional system has approved the separation of powers to prevent the possibility of abuse of executive power. The same rules should apply to the fourth power, that is, the media, including social media. The greatest possible dispersion of the media means a lower risk of the abuse of media power when selecting and controlling other types of power.

Social media ownership concentration is a threat to the basic function of the media, which serves the public interest, because the greater the concentration of the media, the less the possibility for citizens to get a broad range of information. Reducing the number of independent media sources decreases the number of views that the media provides to the public.

It is evident that there is a clear connection between the reduction of media pluralism and the reduction of media coverage of matters of public interest. There also exists a connection between a lower quality of media coverage and public opinion levels and government policy. A lower level of media pluralism impairs the respect of human rights by the executive power. If there is little media competition, the media are less likely to report human rights violations and, on the other hand, executive power is more likely to implement repressive policies.

Generally speaking, in a non-competitive media environment, consumers are less able to see whether the goods offered to them have value. Additionally, they have a reduced ability to assess the quality of information they receive due to the reduction of media system pluralism.[32] The connection between the free competition of ideas on social media and the quality of the information they provide is not strictly linear. This relationship can be described using a curve, such that state-controlled media only provide information in favor of the government.

32 Rozehnal, 2015.

Another extreme is the full commercialization of social media, which would lead to content being provided solely on the principles of profitability. State-controlled media would only provide one-sided information, and the content would be propagandistic. However, if media content was only provided on a commercial basis, and additionally by an oligopoly or monopoly, the media would only try to maximize profits and provide information of poor quality.

The theoretical justification for social media diversification is based on normative democratic theory, according to which political power should be distributed equally. In practice, this means tone person, one vote. Social media power should be distributed in a similar way.

Of course, this is literally impossible, but the concept of the media should be close to such a conception. Another reason for media diversification is Marquis de Condorcet's theorem, called Condorcet's jury theorem. According to this theorem, the more people who make decisions, the more likely they are to make a better choice.[33]

In the context of social media and the electorate, we can assume that the more voters know about a matter through a sufficient number of independent media sources, the more probable it is that a good decision on the matter will be made, especially in elections. Moreover, if there is enough media choice, consumers have the opportunity to check the information provided by different sources.

Efforts to harmonize the regulation of media pluralism and ownership in the EU failed, showing how politically sensitive the matter was. Social media cannot be conducive to democracy without a plurality of media voices and opinions. Pluralism is a basic general rule of European media policy.

Therefore, media power concentration is considered an adverse phenomenon. The European Union upholds media pluralism as the essential pillar of the right to information and freedom of expression guaranteed by Article 11 of the Charter of Fundamental Rights of the European Union.

7. The Product of Social Media as a Public Good

Production of a public good, such as social media production, has a positive effect on society beyond the sphere of a person who consumes the public good. The existence of healthy social media free from government or owner control is a public good, which is a benefit to the majority, even of those who are not interested in the news, do not use social media, and are passive in political matters. The capitalist market usually does not work well in creating a public good because profit maximization corporations cannot give enough weight to positive, universal benefits.

33 Rozehnal, 2020.

Public goods must thus be supported by the public. I realize the problematic nature of public support. Many such discussions have been conducted in the past. In 1792, there was a debate in the American Congress that resulted in paying postage to publishers by the state to deliver newspapers to readers. This was the federal government's largest expense for a long time. Similarly, in the 1970s, intensive preparations were made in Germany to start publishing public newspapers. The other option is to introduce restrictions on media concentration, which might be the first step in restoring the freedom of the press.

Is such a process non-conservative or illiberal? However, what is conservative in the concentration of power—political, corporate, media, and cultural? This is the opposite of both conservatism and liberalism. Media power diffusion encourages citizens to become involved in public life; it is a manifestation of democracy.

It is reasonable to be skeptical about whether the free market itself can offer an optimal amount of information produced by a sufficient number of ideas on social media that are necessary for good corporate governance. Similarly, an economic preference for independent social media is unrealistic. Probably far more realistic is the development of Internet-independent media that can offer a variety of perspectives spread across the web, especially across websites that are run directly by journalists, which is probably the task of the next generation of journalists.

As no boundaries have been set for technology corporations, digital monopolies have emerged, which, by removing posts, blocking users, and using algorithms for recommending content, can affect election results and thus the state of democracy in a country. Social media, especially those with a large number of users, is not only a business but also has a social and political role, like all media.

Unfortunately, there is a permanent conflict between the right of a social network operator to act opportunistically in response to various market conditions and the concept of the human right to free speech. Social media already has such an impact that it can influence political development in a country.[34] This development is then conditioned by the existing political and regulatory institutions in the country. If these institutions fail, digital companies will gain uncontrollable, and therefore abusive, influence and power. If a social media operator has political interests in addition to commercial ones, it can easily gain control of society.[35]

These corporations have created their own rules for removing posts and users themselves, but these rules have nothing to do with the systems of law of the countries in which the users are citizens. They were adopted without public debate and transparency, and they do not allow any remedies. They derive their legitimacy only from their ownership rights, that is, from the right of the relevant platform operator. Various surveys show that they have nothing to do even with the systems of law of

34 Rozehnal, 2015.
35 Rozehnal, 2020.

the countries from which these operators come and to which the acquirers of the platforms subscribe.

We do not know anything about this process. To a large extent, these rules seem to be applied by automatic filters, probably based on keywords and algorithms that are mysterious to us. However, removing content or blocking a platform user cannot be entrusted to computers; it must be a transparent process that is controlled by humans. The online platform operator cannot assume the role of a court, and its decision to remove any content must be reviewed by an independent body. It must then be possible to remove the content only if it is unlawful, and the unlawfulness can only be declared by a court.

This only shows the absolute and uncontrollable, and therefore, easily abused power that digital societies have. The possibility of restricting freedom of speech, which the removal of a post or blocking a user truly is, must be formulated in such a way that this wording is sufficiently precise and predictable to allow citizens to regulate their behavior. If there is an urgent social need for a ban, the ban must be adequate to the legitimate means, and sufficient grounds for intervention must be substantiated.

Several such disputes have already taken place, especially in the United States, although there have not been many. However, the grounds for these decisions are inspiring. For example, the U.S. Federal Court has ruled that an open public official's Facebook page is a public forum and that its founder must not block other users or their posts because of the content of those posts. In the case that I am talking about, it involved a representative of a local government, where FB would delete posts of its Facebook page visitors criticizing the ' management of funds by her colleagues. The lawsuit was based on the fact that this removal from the public forum violates the First Amendment of the Constitution, because it is a place where people should and can express their opinions. The grounds of the judgment further state that it is not possible to prevent people from joining public debates because of their views. The court also explicitly stated that the fact that the website is operated by a business corporation is not a consideration. The right to criticize is at the heart of the First Amendment.

The fact that the founder of the site does not agree with the opinions of other users does not mean that he can silence them. It is just as impossible to silence a person, for example, during his speech in a public park. On the other hand, there is a known case in which a court upheld a blockade of the FB page of the Sikhs for Justice, which is a human rights organization fighting for Sikh independence in the Indian state of Punjab. However, other court decisions have approved this blockade.

Individual countries are too weak to fight large social media platforms, even the world's only superpower, the United States of America. In addition, cyberspace has no borders, while national jurisdictions have them. However, if the European Union properly grabs the opportunity that now lies on the ground, it could be the party

that brings some order to the monopolistic and undemocratic environment of social media.

The argument that social media companies can set rules at their discretion is odd. These corporations provide a service, and each service must be provided on a non-discriminatory basis. Therefore, it is not possible to restrict users simply because they publish a post that is classified as harmful according to the rules of the company providing the digital service.

It is the same as if an electricity distribution company refuses to supply electricity to a house on a street where it otherwise supplies it, pointing out that the occupants of the house are racists who have a negative attitude toward minorities or condemn immigration. Online platforms cannot play by different rules.

While the world is digital, legal regulation is still analog. We regulate political agitation on radio or television, but digital media, which is becoming increasingly important, is still outside our purview. For 20 years, the Internet has evolved with minimal rules to become a truly digital Wild West. Freedom of speech on the Internet must be guaranteed to all, not just to some strong information service providers. Their procedures must be transparent and they must be accountable for their actions. It is time to set some basic rules so that online democracy remains a real democracy.

The purpose should be the creation of a safer and more open digital space centered on values such as freedom, democracy, and respect for the rights of the individual. The huge development of digital services has changed the world. It has changed the way we communicate, our access to information, and the way we buy and use services. European legislation must keep up with this issue.

The digitalization of the world has brought undisputed benefits, but it has also caused some problems, such as the sale of illegal goods and services or the distribution of illegal content, as well as the silencing of so-called harmful views. In addition, online services can be exploited by using manipulative algorithms. This affects our fundamental rights and freedoms. The digitization of society and the economy has caused several major platforms to control an important part of our lives, including the economy and the distribution of information.[36] Social media platforms became rulers of the digital market with the rule-making power of an absolute ruler, creating rules that are unfair to users. All of this should be corrected by the aforementioned law. However, the devil lies in the details, and the result of the regulation will depend greatly on its final form and implementation.

If we do not want to become obedient payers supporting large digital companies and sooner or later following their political preferences, this is definitely a necessary step. What is free about the concentration of power—political, corporate, media, and cultural? This is the exact opposite of freedom and liberalism. On the contrary, the diffusion of power encourages citizens to participate in public life, and it is indeed a

36 Duspiva, 2004.

feature of democracy. Freedom does not consist of a life without rules, but in a life with rules that guarantee equal opportunities for all, respect their rights, and stand on the side of the individual and not powerful corporations. However, it is necessary to be careful about over-regulation.[37]

37 Weiler, 2002.

Bibliography

BURTON, G., JIRÁK, J. (2001) *Úvod do studia médií*. 1st Brno: Barrister & Principal.

CRONE, T. (2002) *Law and the Media*. Oxford: Linacre House.

DRGONEC, J. (2013) *Sloboda prejavu a sloboda po prejave*. 1st Šamorín: Heuréka.

DRGONEC, J. (2008) *Základy masmediálneho práva*. Bratislava: Bratislavská vysoká škola práva.

DUSPIVA, Z. (2004) *Digitalizace jako budoucnost elektronických médií*. Praha: Votobia.

KNAP, K. (2004) *Ochrana osobnosti podle občanského práva*. 4th Praha: Linde.

MCLUHAN, M. (2003) *Understanding Media: The Extensions of Man*. Corte Madera : Gingko Press, Incorporated.

PEMBER, D. (2001) *Mass Media Law*. New York: McGraw-Hill.

RADBRUCH, G. (1999) *Rechtsphilosophie. Studienausgabe*. Heidelberg: Hrsg. R. Dreier, S. Paulson.

ROZEHNAL, A. (2008a) *Media Laws – A Commentary*. Prague: ASPI.

ROZEHNAL, A. (2008b) *Protection of Personality in the Media*. In Bystřický, J. (eds.) *Media, Communication and Culture*. Pilsen: Aleš Čeněk Publishing House.

ROZEHNAL, A. (2011) *Operation of Radio and Television Broadcasting Act. Commentary*. 2nd Updated Edition. *On-Demand Audiovisual Media Services Act. Commentary*. 1st Prague: Wolters Kluwer ČR.

ROZEHNAL, A. (2015) *Media Law*. Pilsen: Aleš Čeněk Publishing House.

ROZEHNAL, A. (2020) *Media Law in the Czech Republic*. Kluwer Law International.

WEILER, P. (2002) *Entertainment, Media, and the Law*. St. Paul: West Group.

ZELEZNY, J. (2001) *Communications Law*. Belmont: Wadsworth, a division of Thomson Learning, Inc.

Chapter VIII

The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism in Slovakia

Gábor Hulkó

1. Introduction

The technological capabilities of Internet communication, the existence (or non-existence) of constitutional foundations for social media, and whether state regulation, self-regulation, and national or global regulations are appropriate for social networks are not yet clear. State action is limited by the jurisdiction system, and in the case of the global self-regulation of service providers, no rule of law guarantees the restriction of fundamental rights. In many cases, such laws are arbitrary.[1]

When considering the state regulation of social media, it is worth distinguishing two problems: *the assessment of disputes and legal liability between users* and *the legal liability of platforms.* In the case of the settlement of disputes between users, in the countries examined in this study, users can sue each other in the same way as in the offline world, or they can conventionally accuse if they suspect that a crime has been committed. The legal procedures remain the same, but the specifics of communication on the social media platform must be considered when investigating an infringement.[2] *The regulation*

1 Klein, 2018, p. 38.
2 Koltay, 2019, p. 14.

Gábor Hulkó (2021) The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism in Slovakia. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 245–276. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_8

of the liability regime for content on social media platforms is another matter and raises different questions: first, the responsibility of social media platforms for user-uploaded content; second, the reaction of social platforms to this uploaded content: whether they ban users' posts and delete (censor) information. In this regard, social media platforms can influence the flow of information at the local or global level; thus, they *de facto* intervene with individuals' freedom of expression and right to information.

2. Regulatory and institutional framework of freedom of expression and censorship in Slovakia

Democratic society and the rule of law guarantee every individual the right to express their views orally, in writing, in print, through images, or otherwise and freely seek, receive, and impart ideas and information, regardless of national borders. *Freedom of expression and the right to information are guaranteed in Art. 26 of the Constitution of the Slovak Republic*,[3] and their limitations and obligations are also stipulated by law. These rights can only be restricted if the measures in a democratic society are necessary to protect the rights and freedoms of others, the security of the state, public order, and public health and morality.[4] Public authorities are obliged to provide information about their activities in the state language in an appropriate manner, while the conditions and manner of implementation are established by law.

The Slovak legal system respects the protection of personal data and provides restrictions on access to or non-disclosure of information, such as possible infringements of intellectual property protection or concerns about decision-making by courts or law enforcement agencies. Restrictions have also been established for other special regulations. *Regarding the conflict and realization of the rights to information and protection of personality, one must be limited in favor of preserving the other.* A special status is acknowledged for public figures and representatives of state power for whom the limits of admissible criticism are extended—the expression of critical opinions about the behavior of certain individuals must be allowed within the enwidened boundaries of freedom of expression.[5]

Freedom of expression guarantees the right of citizens to express their thoughts and opinions, which can only be restricted by law. According to the Slovakian constitutional approach, it is a human right. More *precisely, it is a political right that ensures the dissemination of different political views and allows citizens to influence*

3 *460/1992 Zb. Ústava Slovenskej republiky.* Available at: https://bit.ly/2YNuYU5.

4 Constitution, Art. 26(4).

5 Representatives of state power or public figures must realize that when obtaining this status, the rules also include certain restrictions on their rights to private life, and they may be the subject of wider and sharper criticism in the public interest and the interest of their political opponents.

political developments in the state and participate in public events. Further, it allows protest rallies to take place and uncensored public information on public affairs to be disseminated, as well as the confrontation of the thoughts of an ordinary citizen with the attitudes and opinions of other people. As the freedom of each individual ends where the right of another begins, freedom of speech cannot be abused to interfere with the right of another person. Negative information, even if untrue, can reduce a person's credibility in society and authority in the workplace and disrupt their social relationships. *The provision of false information about events and the abuse of freedom of expression to commit violence are prohibited.*

Everyone is a holder of the right to freedom of expression—not only a natural person but also a legal person, a stateless person, or a group of persons without legal personality (petitions committees, party preparatory committees, and consortia). Every subject to the right to freedom of expression falls under the protection guaranteed by Art. 26 of the Constitution.[6]

The *Constitution of the Slovak Republic*[7] *defines cases in which freedom of expression may be restricted* and sets three basic conditions: a) the restriction of freedom of expression is defined by law[8]; b) a legitimate purpose of the protection of public or individual interest (the protection of the rights and freedoms of others, security of the state, public order, protection of public health and morals); c) the restriction can be considered a measure necessary in a democratic society. This restriction is possible only within the meaning of Art. 26 (4) of the Slovak Constitution by law, on constitutional grounds.

The fundamental principles of the liability system of social media platforms are based on freedom of expression and constitutional rules concerning access to information. Slovakian regulations do not distinguish between online and offline "forums"—*the medium through which the expression of opinion takes place is irrelevant.* Taking a general approach, several aspects of legal responsibility for expressing opinions can be distinguished on online interfaces. In the case of the *private law aspect* of an infringement, the right of privacy is typically violated: this may involve an interference with some personal data, violation of human dignity, privacy, or defamation. These cases typically comprise disputes between individuals ultimately decided by a court. *From a public law perspective*, we can distinguish between administrative-type violations and related administrative sanctions, and in more serious cases, criminal acts and penalties. Administrative-type infringements in Slovakia typically include personal data protection and conflicts of expression, in which case the data protection authority acts in public proceedings and may impose administrative measures and fines. The subject of these proceedings is the protection of personal data, but it does not exclude the possibility of enforcing damages in a civil law procedure. Similarly, in other administrative

6 Filip, 1998, p. 625.

7 Constitution, Art. 26.

8 For instance, criminal acts involving racist statements, symbols of fascism, or lies with the intention of harming others.

sectors, freedom of expression may conflict with other rules and prohibitions. A further public-law restriction on freedom of expression is the framework established by criminal law. In particular, the current criminal Slovak law categorizes hate speech as other types of criminal acts (incitement against the community, use of an authoritarian symbol, or incitement to violence). Nevertheless, it is possible to differentiate the *legal means of protection in case of abuse of freedom of expression* as follows: a*) civil law protection* (protection of personality, good reputation); b) *criminal protection* ("hate speech"); c) *administrative law protection* (broadcasting and retransmission regulation in media services, press regulation, or regulation on advertisements, consumer protection).

Censorship is also prohibited constitutionally: censorship is forbidden.[9] In the prevailing doctrine in Slovakia, the concept of "censorship" is only relevant in the relationship between state and freedom of speech—the regulation is directed toward the state and its organs (*de iure* censorship). Therefore, this constitutional rule does not apply to actions of private individuals or corporations capable of limiting, banning, or *de facto* censoring the views of others.[10]

From an institutional perspective, there is no state or administrative organ that actively and explicitly supervises the freedom of expression. However, regarding this fundamental right, several public administrations perform subtasks within their sector. These include, in particular, the Office for Personal Data Protection (*Úrad na ochranu osobných údajov*),[11] the Council for Broadcasting and Retransmission (*Rada pre vysielanie a retransmisiu*),[12] and the State Committee for the Supervision of Electoral and Political Party Financing (*Štátna komisia pre voľby a kontrolu financovania politických strán*).[13] Broadly, this also includes the Council of Slovak Radio and Television (*Rada rozhlasu a televízie Slovenska*),[14] which conducts public service media oversight. Overall, its task is to guarantee and control the independent operation of public service media and the provision of objective and balanced information.

The *Office of Personal Data Protection* is primarily responsible for state tasks in connection with personal data protection. Freedom of expression in the operation of this office is affected in the context of personal and protected data. The *Council for Broadcasting and Retransmission* performs certain state tasks in the field of radio and television. Its main mission is to promote public interest in the exercise of the right to information, freedom of expression, cultural values, and access to education in the sector, in particular licensing, supervision, sanctioning, and individual administrative tasks. The Council does not independently monitor the exercise of freedom of expression but may generally examine it (in the context of some

9 Constitution, Art. 26(3): Censorship is banned.
10 See below for more detailed elaboration on this matter.
11 Available at: https://dataprotection.gov.sk/uoou/.
12 Available at: http://www.rvr.sk/.
13 Available at: https://www.minv.sk/?statnakomisia.
14 Available at: https://www.rtvs.org/rada-rtvs/o-rade-rtvs.

other objective). The *State Committee for the Supervision of Electoral and Political Party Financing*—responsible for overseeing the financing of elections and political parties—is primarily involved in the financial oversight of party operations and the transparency of election campaigns. It may control freedom of expression only tangentially in its activities.

3. Constitutional and legal sources of the regulation of freedom of speech

Art. 26 of the Constitution of the Slovak Republic[15] provides a constitutional framework for freedom of expression (*sloboda prejavu*) and the right to access information (*právo na informácie*). Pursuant to Art. 26 (1), freedom of expression and the right to information are guaranteed in the territory of the Slovak Republic, enjoying constitutional protection. Accordingly, under para. 2, everyone has the right to express their views orally, in writing, in the press, through images, or otherwise and freely seek, receive, and impart ideas and information, regardless of frontiers[16]. Both freedom of expression and the right to information may be restricted, as stated above. The prohibition of censorship is stated in Art. 26 (3) of the Constitution.

Provisions more narrowly or broadly related to freedom of expression are contained in several pieces of legislation. These include, in particular, the following: provisions of the Civil Code (*zákon č. 40/1964 Zb. Občiansky zákonník*)[17] on personal rights, general liability and compensation; facts of the Criminal Code (*zákon č. 300/2005 Z. z. trestný zákon*)[18] concerning violations of the rules of community coexistence (e.g., incitement against a community, violence against a member of a community); Act no. 22/2004 on Electronic Commerce (*zákon č. 22/2004 Z.z. o elektronickom obchode*)[19]; Act no. 211/2000 on Free Access to Information (*zákon č. 211/2000 Z. z. o slobodnom prístupe k informáciám*)[20]; Act no. 18/2018 on Personal Data Protection (*zákon č. 18/2018 Z. z. o ochrane osobných údajov*)[21]; Act no. 185/2015

15 *460/1992 Zb. Ústava Slovenskej republiky*. Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/1992/460/ (Accessed 31 May 2021).

16 Constitution, Art. 26(2): Everyone has the right to express their opinion in words, writing, print, images, or otherwise and seek, receive, and disseminate ideas and information freely, regardless of the state borders. No approval process shall be required for press publishing. Entrepreneurial activity in the field of radio and television broadcasting may be subject to permission from the State. The conditions shall be laid down by a law.

17 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/1964/40/20191201.

18 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2005/300/20210101.

19 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2004/22/.

20 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2000/211/20210101.

21 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2018/18/20190901.

on Authorship (*zákon č.* 185/2015 *Zb. Autorský zákon*)[22]; Act No. 308/2000 on Broadcasting and Retransmission (*zákon č. 308/2000 Zb. o vysielaní a retransmisii*)[23]; Act no. 167/2008 on Periodicals and News Agencies (*zákon č. 167/2008 Zb. o periodickej tlači a agentúrnom spravodajstve*)[24]; Act no. 372/1990 on Misdemeanors (*zákon č. 372/1990 Zb o priestupkoch*)[25]; Act No. 351/2011 on Electronic Communication (*zákon č. 351/2011 Z. z. o elektronických komunikáciách*).[26] Another feature of the general legal framework is that the Slovak legislator has not yet implemented the 2018 amendments[27] to the AVMS Directive, which are listed in the legislative plan of the government.[28]

4. Legal sources and general rules of social media platforms

Social media platforms are *not specifically regulated* in the Slovak legal system. The only current regulation that has some direct relevance to social media liability (social media vs. state relation) is based on the abovementioned *Act No. 22 of 2004 on electronic commerce* (hereinafter referred to as the e-Services Act). This act was passed to transpose the rules of the e-Commerce Directive[29] into national law—*it is not targeted at regulating social media platforms specifically,* but it can theoretically also be applied to them.

Furthermore, Slovak legislation does not define a special concept of illegality or infringement, either in relation to e-services or social media. Accordingly, an infringement is considered to be any infringement under Slovak law. In the case of the removal of infringing content and the infringement suffered online, an individual can, as a general rule, *seek out a court*. In some branches, such as the protection of personal data and the protection of copyright, there is an administrative supervisory body. *Therefore, administrative intervention is also conceivable under sectoral legislation*. *The investigating authorities may act* on suspicion of a criminal offense. The regulation does not differentiate between infringements committed in

22 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2015/185/.
23 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2000/308/.
24 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2008/167/20191101.html.
25 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/1990/372/20210501.
26 Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2011/351/20210801.
27 Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities.
28 Available at: https://www.slov-lex.sk/legislativne-procesy/-/SK/dokumenty/LP-2020-622.
29 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

the "online" and "offline" space; thus, the nature of the medium is irrelevant to the availability of legal means. According to Slovak regulations, the individual layers of user-state-provider relationships in the online space can vary, as described below:

Relation type	**Example**	**Possible action/consequence**
a) user vs. user	personal right protection; copyright infringement; personal data protection	court and/or administrative action
b) user vs. state	hate speech or similar unlawful act (criminal acts, administrative offences)	police investigation, court and/or administrative action
c) user vs. provider	removal of users' content; alleged censorship	providers' terms and policies or court
d) provider vs. state	passive provider (e-Services Act); providers that commit unlawful acts	administrative or court action

Court procedures can occur in *a) user vs. user* disputes on social media platforms; however, for special violations (such as data protection and copyright violations), administrative procedures can be initiated, and administrative legal consequences are determined. The *b) user vs. state* relationship is mostly relevant in cases of serious breaches of the law (e.g., criminal acts), for which the police organs undertake the necessary means to stop such behavior. This can be followed by criminal court procedures. The *c) user vs. provider* relationship is considered a private contract between private individuals under Slovak law. Therefore, for unsolvable disputes under the terms of service (such as the removal or banning of users or their content and restricting users' information), the plaintiff can turn to court to resolve the issue. The liability of the *d) provider to the state* describes the responsibility for online content relayed and/or displayed by the provider. In the latter case, the service provider's liability for content is significantly limited and practically excluded. According to the law, the *service provider is not responsible for the transmitted information if the provision of the service comprises only the transmission of information in the electronic communications network or the provision of access to the electronic communications network*. Simultaneously, the service provider should not have a) initiated the transmission, b) selected the recipient of the information, or (c) compiled or modified the information. Furthermore, the service provider *shall not be liable* for information stored in the memory of the electronic devices used for information retrieval at the request of the user, *provided that the service provider is not aware of the illegal content of the stored information and takes immediate action to put an end to the user's unlawful conduct.* For such information, the service provider shall be liable only if the user acts on its instructions. In summary, *the service provider is responsible for the content it stores or transmits if*

a) it has become aware of its illegality and has not acted against it or if b) it has had a significant influence on the compilation of the information. With these provisions, the legislator transposes Art. 14 of the e-Commerce Directive into Slovak law with virtually no substantive changes.

Thus, it would be considerably difficult to hold the largest group of service providers liable for the information they store or transmit. Exceptions are news portals (online newspapers, magazines), online radios, and television channels—all service providers who produce their own information or news or have a significant influence on it. The forums of such providers are moderated posts that violate rights or public morality are removed, following the rulings of the European Court of Human Rights on October 10, 2013, in Delfi AS v. Estonia's[30] verdict.

Furthermore, *service providers have no obligations to monitor users' content, and the regulations explicitly prohibit the service provider from searching the data users transmit or save* without their consent. Nevertheless, if the provider becomes aware of the illegality of such information, it shall remove or at least prevent access to it, and the court may order the service provider to remove the information even if the service provider is unaware of its illegality. Thus, the search for user information is generally excluded, so the service provider *has no obligation to actively search for content (content tracking).*

Apart from political statements and newspaper reports,[31] there is *no common legal or scientific position* on the *obligation for social media platforms to intervene against illegal users' content,* nor *on the removal or banning of user-generated content,* except the e-Services Act Slovak. The literature cites the court case of *Stacho v. Klub Strážov,*[32] in which a comment on a website that violated the human dignity of a specific individual (not an article but a reader's comment) was disputed.

5. Content censorship in social networks in Slovakia

5.1. General rules on censorship

As stated above, the Constitution of the Slovak Republic prohibits censorship but does not define it. Likewise, the legal definition does not exist in any valid law. The only legal definition of the term was provided by the previous press regulation, Act no. 81/1966 on periodicals and other mass media (*zákon č. 81/1966 Zb. o periodickej tlači a o ostatných hromadných informačných prostriedkoch*): *Censorship refers to any intervention by state authorities against freedom of speech and image and their*

30 Delfi AS v. Estonia. Available at: https://bit.ly/2XmB2Te.
31 See: https://bit.ly/3zb9IUX and https://bit.ly/2XkaNw8.
32 Husovec, 2012.

dissemination by mass media. This is without prejudice to the powers of the prosecutor and court.[33] As this is the only legal definition of the term included in the Slovak legal system, it remains relevant to date,[34] although it was replaced by the current regulations in 2008.

Without a *definitionem legis,* various attempts have been made to define. According to the prevailing theoretical approach, *censorship is an official examination of everything intended for publication (especially the press), considering state, political, and moral interests, including the possibility of an official ban on publication*[35]. The opposite side of censorship should also be considered, such as in cases in which the court in a personal protection dispute orders the defendant to refrain from making statements in the future that violate the plaintiff's right to protection of personality. Censorship is not a publisher's obligation to withdraw a book the content of which infringes on the personality or copyright of another person. In all such cases, these are measures taken by public authorities based on proceedings initiated in relation to specific subjective rights or public interests that are directly endangered[36]. *Inadmissible censorship includes institutional, preliminary (preventive), and subsequent censorship*.[37]

Furthermore, *self-censorship cannot be subsumed under the accepted definition of censorship in the sense of the Constitution*. This is also a relevant issue, as self-censorship refers not only to individuals' self-restraint in their writing or speech but also to an editor's refusal to publish anything in a newspaper or magazine or a publisher's requirement to edit a book not in conflict with copyright law.[38] Excluding this type of censorship "sweeps it under the rug," pretending there is no problem while it erodes freedom of speech.[39]

Most authors define relevant *conceptual features of censorship's public power nature—prohibition of censorship is addressed exclusively to the state*. Interference with freedom of expression by private individuals—while not necessarily less threatening than interference by public authorities—cannot be classified as censorship.[40] While most authors take this definition for granted, other approaches to this issue underline that the concept of "censorship addressed exclusively to the state" is outdated and should be revised. Forms of communication have changed and

33 See § 17 para. 2 of Act no. 81/1966. This regulation was in force only between June 28, 1968, and September 25, 1968. After September 1968, this para. was abolished and the definition never used again. See: https://bit.ly/3hVf0xQ.

34 Drgonec, 2015, pp. 61–79.

35 Bartoň, 2002, pp. 21–22.

36 Moravec, 2013, p. 34.

37 Pavlíček et al., 1999, p. 182.

38 Ibid.

39 In dealing correctly with the issue of freedom of expression, self-censorship may not be a sufficiently intense issue interfering with freedom of expression. It does not affect the mass media in the institutional sense and cannot be entirely neglected, as it represents a problem for freedom of expression at its core. See Drgonec, 2015, pp. 67–68.

40 Drgonec, 2015, p. 64.

developed drastically in the last two decades, and many communication platforms (such as social networks and mass media) no longer fit into the "traditional" concept of censorship. This creates a dangerous possibility that *de facto censorship, which shows all the hallmarks of censorship, will remain outside de jure censorship*—formal protection against censorship will be considerably remote material protection from censorship.[41] The half-a-century-old legal definition of censorship appears as if it is from another world because it comes from a completely differently organized Czechoslovak socialist state and society. Therefore, it should not automatically be used as the definition of censorship, which is the subject of Art. 26 par. 3 of the Constitution.[42] Censorship refers to the control of the content of disseminated information, the control of information sources, and institutionalization in the form of a submission and authorization obligation with the possibility of a power ban. This power is available to public authorities, the owners of mass media, and the employers of journalists as those affected by the speech. The ban on censorship is aimed at providing protection against censorship to any expression under constitutional protection.

5.2. Censorship in the decisions of the Constitutional Court

The interpretation of the constitutional provision on censorship is rather scarce in the legal practice of the Constitutional Court of the Slovak Republic. The latest decision considering the interpretation of Art. 26 par. 3 of the Constitution was Decision IV. ÚS 307/2014,[43] in which the Constitutional Court states that "Censorship (direct censorship) in the constitutional sense means mainly the *politically motivated intervention of public authorities in the freedom of expression* of the subject concerned. This comprises assessing the content of opinions, thoughts, ideas, facts, and their form of dissemination and representation intended by the subject, publisher, etc., in the future (ex-ante control) or which have already been made available to the public (ex-post control) to change or completely negate these views,

41 Arguably, "situations in which freedom of expression could be denied to an individual are constitutionally unacceptable because most other individuals can express an opinion without censorship. Everyone is subject to freedom of expression; therefore, everyone has the right to protection from censorship. The prohibition of censorship confers protection to every holder of the right to freedom of expression and the right to information. The prohibition imposed in the Constitution, protection against censorship, was granted in the highest available legal force. The constitutional ban on censorship is absolute and applies to all addressees of the ban. The Constitution does not grant an exception to anyone, nor does it exclude anyone from the circle of legal entities obliged to respect the prohibition of censorship. The strength of the traditions formed and formulated by the previous legal definition leads to a restrictive interpretation of the addressees of the ban on censorship, to their identification with the state authorities. Thus, the prohibition of censorship must be interpreted as one not addressed exclusively by a public authority but to any subject of law in a position of power." See Drgonec, 2015, pp. 61–79.

42 Drgonec, 2015, p. 62.

43 See: https://bit.ly/3nsuRrq.

thoughts, ideas, or facts or their form of dissemination and display, mainly for political reasons. *The nature of direct censorship can also impact freedom of expression comprising a ban on the dissemination or an additional ban on the dissemination of certain types of information that have been disseminated without restriction in the past* (e.g., a ban or additional ban on publishing, a ban or additional ban on the media, a ban on public publication). The reasons can be related to the subject or the content of disseminated opinions or ideas the subject concerned has in the past or at the time intended or disseminated unless it is prohibited for reasons justifiable by the Constitution."[44]

The Constitutional Court takes the same essential approach as the prevailing doctrinal view. Furthermore, in the aforementioned decision, the court only considers the definition and context of direct censorship and does not elaborate on other types of censorship. Despite the relevance of freedom of expression and its relationship with censorship in democratic societies and considering the consequences of censorship for the availability of freedom of speech, accessible legal literature and court decisions have focused on identifying the scope and content of the ban on censorship under Art. 26 par. 3 of the Constitution only marginally.

5.3. Censorship in social media

As the main legal doctrine of the ban on censorship in general aspects considers as "censorship" only actions with a nature of public power, there is no unified view on censorship of social media platforms. Slovak law only stipulates tangential rules on freedom of expression on the Internet and social media in certain provisions of the e-Services Act (as stated above), which may have a certain degree of applicability in this area. However, notably, this law was adopted to transpose the rules of the e-Commerce Directive into national law. This legislation is therefore not primarily intended to regulate online expression and social media but transposes the general liability of hosting providers laid down therein in accordance with the European Directive on Electronic Commerce. There is currently no legislative intention to regulate freedom of expression online or content removal of social media platforms.

The material scope of the e-Services Act is the information society services (*služby informačnej spoločnosti*). According to the regulations, such a service is any information society service provided *remotely* (e.g., the service is provided without the simultaneous presence of the parties), *electronically* (e.g., the service is sent from the point of origin and received at its destination entirely by wire, radio, optical, or other electromagnetic means), *usually for a fee* and *at the individual request of*

44 Ibid.

the recipient.[45] When examining the Slovak legislation, we can consider the general terms and conditions of service providers—bilateral private law agreements—to be of a "soft law" nature. The general approach to such contractual terms is that service providers reserve the right to remove infringing content, and the user can make court petitions to seek remedies.

In summary, as there is no regulation of social media platforms in Slovakia, *the system of public liability for content control by social media platforms is not stipulated in the Slovak legal system with regard to alleged or actual censorship*. Prohibiting the sharing of information or its removal from a particular medium may be ordered by a court, but service providers must remove the infringing content themselves. *The relationship between social media and the user is interpreted by Slovak law as a private law contract* within the framework of which the user consents to the service provider to remove certain (infringing) content. Thus, *if information (entry, comment) is deleted by the service provider, this can be challenged in court, but the Slovak legal system* does not provide other guarantees. In practice, such cases (in which the user files a lawsuit against social media platforms) do not occur.

6. "Fake News" and the influence of digital platforms and social networking on the guarantees of freedom of speech and truthfulness of information in Slovakia

There are currently no valid regulations of fake news in Slovakia. There were several instances in which the Slovak Police Force fought false information or misinformation about the latest COVID-19 measures in the country. This primarily included a heightened presence on the "official" social media channels of the force, in which constant and fact-checked information of the population was ensured. In parallel, false information and hoaxes were monitored, and the civil population was constantly informed about false news. There were also several cases in which interventions were taken against users from Slovakia for spreading false information on social media concerning the spread of the virus or the pandemic situation in general. These interventions were conducted among individuals who could be identified and tracked in Slovakia. These actions were undertaken in the context of the pandemic, and there are no known cases in which a social media platform

45 According to Slovak legislation, the provision of information society services in Slovakia does not require a permit or registration (notification). However, the provision of the service may be restricted if the service provider violates the requirements of state security, public order, public health, or environmental and consumer protection.

or other service provider was encouraged to ban users' posts or users on behalf of the police.

However, the Slovak government fully realizes that advances in information technology have provided citizens with access to extensive information and the creation of information. Much of such information is often misleading and/or untrue. The massive spread of various misinformation is increasing as one of the means of the "hybrid war." Such information operations are not new, but with the emergence of new platforms and more effective dissemination techniques, their impact on state security is rapidly increasing. As government documents state, *the term disinformation has not yet been codified in the Slovak legal system*. Simultaneously, disinformation can be considered part of a broader process called information operations in terms of information manipulation. The public is also exposed to the growing dissemination of dangerous rumors, misleading information, and conspiracy theories that can endanger human health, harm the cohesion of society, or lead to public violence and social unrest. *In addition to the targeted dissemination of potentially harmful information, information operations may involve the collection of sensitive data*, encourage people to take action (violent or non-violent), and openly or covertly promote a party or state.[46] Information operations have become the most frequently used hybrid indicator—not only of foreign actors—within the hybrid threat.[47]

In December 2018, the European Commission presented an *Action Plan to Combat Disinformation,*[48] the main aim of which was to strengthen existing mechanisms and build new ones to eliminate this dangerous phenomenon, including the use of artificial intelligence. The need for cooperation in the fight against misinformation within the EU is also one of the objectives of the forthcoming *European Democracy Action Plan*.[49] In this context, the results of the Eurobarometer[50] are alarming and present an argument in favor of addressing the issue, as 83% of respondents described online misinformation as a threat to democracy, a view consistent in all EU countries. At least half of the respondents stated that they encountered disinformation at least once a week, with the most positive answers recorded in Spain, Hungary, Croatia, Poland, France, Greece, and the Slovak Republic.

Official sources imply that *the Slovak government sees information operations as the greatest risk to national security*, as they can be conducted by foreign state and non-state actors (also by domestic actors who sympathize with the attacker). Sophisticated strategies are often taken to influence public debates, deepen the polarization

46 Such a systematic use of information operations is included among the hybrid indicators, which can become hybrid threats.

47 For details, see: https://bit.ly/2XdiAMZ.

48 Available at: https://bit.ly/391iSsf.

49 Available at: https://bit.ly/3EgIeRl.

50 On the final results of the Eurobarometer on fake news and online disinformation, see: https://bit.ly/2YJrp18.

of society, and create a growing group of people who do not trust any official source and are thus more easily manipulated. The result is a more effective intervention in democratic decision-making, the relativization of the country's political leadership, and the weakening of society's confidence in democratic institutions. *The role of the state and its competent components is to create a mechanism to eliminate the impact of disinformation campaigns*, especially through the effective identification of manipulative content and strategic communication.[51]

According to these findings, *the Slovak government's official perspective is that the state must strengthen its means and capacities for resilience to information operations* and cooperate with experts from the public and private sectors *to detect and analyze false information*. In its *program statement, the Government of the Slovak Republic*[52]*undertook to prepare an action plan for the coordination of the fight against hybrid threats and the spread of disinformation and build adequate central capacities for its implementation*. However, these steps must be in accordance with human rights legislation and must never weaken freedom of speech and the unrestricted access to information, which are basic human freedoms guaranteed by the Constitution of the Slovak Republic.

Realizing state security risks, the Slovak government has prepared a novelization of *Act no. 69/2018 on cyber security*[53] and an *administrative action plan*[54] (a coordinated mechanism of the Slovak Republic's resilience to information operations). In particular, the latter provides detailed insights into measures that the government plans to implement in this field.

6.1. The concept of state intervention against "Fake News" in Slovakia

6.1.1. General concepts regarding harmful information

In the context of the dissemination of potentially harmful information, there are numerous *elements of information operations*—activities or methods of implementation (hereinafter referred to as the "EIO"). The most well-known and most frequently used EIOs include the following:

a) *False reports* (fake news) comprise information that intentionally mimics the format of a news or other journalistic product, with its creators deliberately misleading their audiences by distorting reality.
b) *Hoax* includes deceptions, jokes, and virally extended alarm messages. They usually have three features: urgency, reference to illusory authority (such as police sources and scientific results), and requests for dissemination. A common intention is to cause fear or anxiety.

51 Ibid.
52 See: https://bit.ly/3htezKS.
53 See: https://bit.ly/395ao3C.
54 See: https://bit.ly/3AfltuY.

c) *Propaganda* includes information, ideas, opinions, or visual materials created and distributed to influence people's opinions. Propaganda is based not only on half-truths or untruths but also on facts, but it is always biased toward promoting a certain party or opinion. The intent is to induce objectivity despite the one-sidedness of the narrative, the aim of which is to convince and not inform.
d) *Conspiracy theory* explains an event or set of circumstances as a result of a secret conspiracy, usually by a small, powerful group of people. Such a group is usually the government, representatives of secret societies, organizations, or intelligence services, one or more cooperating companies or representatives of states, nations, or religions, or even extraterrestrial civilizations. Conspiracy theories reject the generally accepted explanations of these events.
e) *Parody and satire*, in the context of information operations, are used to disseminate misleading information aggressively or ridicule or criticize a goal (such as a person, group of people, or opinion) that goes beyond the ordinary framework of this genre.
f) *Disinformation* refers to false or manipulated information intentionally disseminated to mislead and harm. Disinformation can be false or manipulated texts, images, videos, or sound, and used to support conspiracies, spread doubts, and discredit true information or individuals and organizations. True information can also be classified as misinformation if presented in a manipulative manner. Misinformation does not include unintentional errors in news, satire, parody, or one-sided reports and comments clearly marked as such.
g) *Malinformation*[55] is based on reality and is intentionally disseminated to harm a person, organization, or state (e.g., leaked information, hate speech, or harassment).[56]

6.1.2. State aims and institutional provisions

The main aim of fighting harmful information on social media platforms is *to reduce and possibly eliminate the space and opportunities* for false and misleading information or news in all areas of public power and achieve society-wide awareness. Thus, it increases public confidence in public authorities, *increasing media literacy* and promoting an information source for objective journalism *to promote more active cooperation and information exchanges.*

55 Malinformation differs from misinformation, which is erroneous or false information spread unknowingly and without intent to harm. Therefore, it is not considered an element of information operations.

56 Council of Europe (2017) Report on Information Disorder: Toward an interdisciplinary framework for research and policy making, DGI(2017)09. Available at: https://bit.ly/2XdHkVl.

The individual state administration bodies of the Slovak Republic have planned a coordinated complex approach at both the vertical and horizontal levels of government and through intensive exchanges of information. The hypothetical goal of the regulation and administrative actions is to establish a consistently well-informed public, for which the government and all organizations and bodies in the public sector are responsible.

Preventive and directly performed activities in the Slovak Republic are planned to be guaranteed by the Government of the Slovak Republic and individual central state administration bodies. The *Situation Center of the Slovak Republic* (hereinafter referred to as "SITCEN")[57]—organizationally integrated into the structure of the Government Office of the Slovak Republic[58] (hereinafter referred to as "Central Office")—will have a specific position in the analysis of the identified elements of information operations (EIO). As part of the institutional system, the *National Security Analysis Center* (hereinafter referred to as "NSAC")[59]—a part of the Slovak Information Service[60]—will play an important role in this analysis, using input from the participating ministries. If it is found that the EIO meets the elements of the factual nature of the crime, the procedure will be left to the law enforcement authority (police organs and prosecutors' offices). Entities operating in the non-governmental sector are also significant in the prevention and identification of EIOs. The state will create a scheme for their involvement and financial support.

According to the administrative action plan, the main *SITCEN* tasks will be as follows: *a)* publish ongoing EIOs of a worrying, high, and critical level of influence and confront it with relevant facts, in consultation with the NSAC; *b)* provide official, comprehensive relevant information on EIOs; *c)* process, analyze, and evaluate EIOs; *d)* use designated software to work with information, obtaining and collecting EIOs, creating analyses, and advancing acquired EIOs; *e)* cooperate with non-governmental organizations and other entities to strengthen the prevention of EIOs; *f)* cooperate with foreign partners to identify possible international cases; *g)* maintain a database of assigned EIOs that may be useful in formulating media outcomes; *h)* cooperate in the development, updating, and use of disinformation software and provide support for public authorities involved in the use of the software; *i)* propose appropriate measures and guidelines to eliminate the spread of EIOs; *j)* contribute to raising awareness of the harmful effects of EIOs and their prevention; *k)* support practical training and education in the field; *l)* to organize conferences to evaluate EIOs over the past year, in the context of prevention, in cooperation with the academic and scientific community and non-governmental sector; *m)* cooperate with the

57 See: https://bit.ly/2VFUVUw.

58 More precisely, it is part of the Office of the Security Council of the Slovak Republic, an organizational part of the Government Office of the Slovak Republic. See: https://www.vlada.gov.sk//bezpecnostna-rada-sr/.

59 See: https://www.sis.gov.sk/o-nas/nbac.html.

60 See: https://www.sis.gov.sk/about-us/introduction.html.

media through consulting and training to keep it informed to eliminate or minimize the spread of EIOs.

The work of the SITCEN will be supported by *NSAC*, and both organizations will coordinate their activities in the field. The main competences of NSAC (which is a part of the intelligence services, as stated above) will be to *a)* cooperate with SITCEN to analyze EIOs, *b)* deliver an opinion according to the level of influence of the EIO, and *c)* in the case of critical EIOs, decide on the course of action.

The proposed material also defines tasks for other *organs of central administration*: a) search and assess EIOs in their area of responsibility manually, analytically, or through specially designed search software; b) prepare a description of the situation and identify the level of the EIO's influence; c) take a position on identified EIOs; d) forward all relevant information to the EIO to SITCEN; e) provide, within their respective spheres of competence, cooperation and additional information on the transferred EIO for SITCEN and NSAC; f) use dedicated, unified software to work with information, obtain and collect EIOs, create analyses, and forward acquired EISs to SITCEN and NSAC; g) use the data from the analyses of their own EIOs and from the outputs of the SITCEN to improve their strategic communication as a basic means of resilience to EIOs.

This institutional framework is supplemented by cooperation with non-governmental organizations, in which *selected non-governmental organizations* will be involved in the possibility of searching for EIOs (also possibly with the use of a designated software for this purpose, for gathering, analyzing, and forwarding EIOs to competent organs) and conducting educational activities.

6.1.3. EIO assessment criteria and state response

In assessing the level of impact, the assessor shall consider the following criteria: *a) the potential to cause harm* (manipulation, polarization of society, human health, economic damage, the rule of law, the credibility of the state); *b) the existence of the potential to provoke action* (non-violent, violent, mass unrest); *c) the size of the group that could be affected* (individual, small group, large group, whole population); *d) originator of the EIO* (individual, group, risk group, non-governmental organization, state organization, state representative); *e) the significance of the influence* of the status of the addressee and *the potential for amplification* (ordinary citizen, member of the risk group, generally recognized personality, civil servant, public prosecutor); f) the degree of probability of influencing the addressee (EIO content quality – *ability to convince the addressee*); g) *credibility of the EIO*; h) *coordination of the dissemination* (unorganized/organized); *i) the EIO's channel in terms of persuasiveness* (oral, social networks, website, print medium, audio visual medium); *j) the disseminator* (individual, group, risk group, non-governmental organization, state organization, state representative); *k) geographical source* (foreign, domestic); *l) characteristics of the conduct showing signs of crime* (defamation, dissemination of an alarm message, incitement to racial, religious or other intolerance); *m) existence*

of neutralization mechanisms (there is/is not a possibility to take countermeasures); *n) other significant circumstances* (timing, concurrence with other elements of hybrid threats, etc.).

Based on these assessment criteria, *the EIO's impact level is defined* based on complex, quantitative, and qualitative analyses. After determining the level of impact, it is necessary for the competent central state administration body to select an adequate response and implement it—*the nature of the response should correspond to the specified level of impact*. As a general rule, the response is implemented by the organ responsible for the sector administration (e.g., in the case of false information about environmental issues, the Ministry of Environment should act; in the case of fake news about public health questions, the Ministry of Healthcare is the competent authority). Depending on the individual characteristics of the assessed EIO, the possible responses fall under the following categories:

a) Negligible influence: There is only a remote possibility that the EIO will have some consequence; there could be an unintentional error in communication or a misunderstanding. While the error can be eliminated, the harmful information cannot trigger action, and the impact can be refuted by verified and documented facts. If the relevant organizational unit evaluates the impact as negligible, the reaction will generally not be necessary. If the competent organ has doubts regarding the level of impact, it consults and coordinates with SITCEN. If SITCEN discovers additional facts, it may change the level of influence.

b) Worrying impact: There is a likelihood of an adverse consequence or the creation of space for the spread of EIO; there may be a risk of harm to the credibility and/or health of an individual or group, violations of law; usually there is unorganized coordination of EIO. In this case, the EIO refers to the SITCEN together with an analysis, the determined level of impact, a description of the situation, and the method of response. SITCEN is obligated to register the EIO in its database, evaluates it, and subsequently forwards all connected information and a description of the situation to the NSAC, which constructs its opinion. Subsequently, the opinion is forwarded through SITCEN back to the competent administrative unit. If the NSAC identifies a different level of impact, its response corresponds to the specified level of impact.

c) High impact: There is a high probability of an adverse event with an impact on the credibility of state bodies, organizations, threats to the health of a group of persons, and threats to the seriousness of a group of persons. The EIO has a high potential to trigger an action, and organized coordination of EIO dissemination was indicated. In the case of high impact, the process of reaction is the same, as in the case of a worrying impact.

d) Critical impact: There is a considerably high probability of an adverse event, a significant threat to the credibility of state institutions and their representatives, the security of the state, significant strategic interests of the state, the

existence of serious damage to the health of a group of people or their lives, high economic damage, endangered sovereignty, territorial integrity, the principles of democracy, and the rule of law. The EIO impacts the whole population with extremely high potential to trigger action; there is highly organized coordination of EIO dissemination. The EIO is caused by a state representative or a state institution; there is excessive room for uncontrollable dissemination. In the case of a critical impact, the EIO refers to the SITCEN from the competent administrative unit, together with an analysis, the determined level of impact, a description of the situation, and the method of response. In this case, the SITCEN is obligated to send the EIO to the NSAC for subsequent analysis, which will consider the need to take measures or convene the NSAC Council, which shall decide on further action.

6.2. The use of social networks in Slovakia[61]

Social networks have become a crucial phenomenon that significantly affects the entire Slovak society. In Slovakia, 86% of the population uses a social network at least once a month, and 61% of people use it daily. The use of social networks is one of the most common activities performed by Slovaks on the Internet.

Facebook is the most widely used social network in Slovakia. At least once a month, Facebook is used by up to 76% of the population of Slovakia, while daily usage is at 55%. Facebook is slightly more popular among younger people; with increasing age, the intensity of its use decreases. Even teenagers who also use many other networks use Facebook daily. Furthermore, Facebook is used by state organs for communication purposes.

Regarding other social media platforms, *YouTube* has versatile uses, although it can be used comfortably without creating one's own account and without using the "social" dimension. Nevertheless, YouTube is used by 78% of Slovaks at least once a month, with 31% of the population using it at least once a day. It its thus the most watched provider of video content in general, even compared to television broadcasting. *Instagram* has become the third most widely used network. Although its core use base is composed of the youngest age groups, Instagram has managed to bridge the generational barrier. At least once a month, Instagram is used by up to 42% of the population, and a fifth of the population (22%) use it daily. Instagram is used by up to 80% of those aged below 26 but only about 10% of those age above 60. Once the most popular social network in Slovakia, *Pokec*, has a lower but stable userbase. Pokec is still used monthly by 19% of the Slovak population, and less than 9% of the population logs onto Pokec daily. It is most often used by people aged 27–40. In addition to large social networks, narrower networks have been

61 This sub-chapter is based on *Koľko Slovákov je na sociálnych sieťach?* (March 2021). Available at: https://bit.ly/3k4V5y7. As well as on *Králi sociálnych sietí na Slovensku: Facebook, YouTube a Instagram* (May 2015). Available at: https://bit.ly/3tzAW6s.

identified in the Slovak market. *Pinterest* has a high penetration (just over 20% of monthly users), although only a small proportion of this are core regular users. *Tik-Tok* also has a growing relevance; Tik-Tok users are often children below 15 years of age (and therefore are not included in the survey). Therefore, the number of real users in the whole population is probably higher than that indicated above, although the use of this social network is gradually reaching higher age categories. At least once a month, 13% of the population of Slovakia use it daily, at approximately 5%. *Snapchat* in Slovakia is currently rather stagnant; it is used by 9% of Slovaks per month, only 3% on a daily basis, while users are exclusively people aged below 26 (the core of the user group are teenagers, similar to Tik-tok). *Twitter* is an interesting case study. While it is used intensively globally (approximately a quarter of the American population has a Twitter account, and the tweets of the former American president, Donald Trump, received global attention practically every day), in Slovakia, Twitter did not catch on. At least once a month, 13% of the population uses it, but less than 3% do so daily. Twitter is thus used extremely passively, and its influence in Slovakia is rather marginal; however, its users are typically better-off people of younger middle age with a higher income and a higher social status. At least once a month, 8% of the population of Slovakia visits the professional social network *LinkedIn*. The audio social network *Clubhouse*, which was given much attention in early 2021, has thus far attracted only a marginal proportion of the Slovak population.[62]

6.3. Legal liability of users and digital media platforms

In questions of legal liability connected to freedom of expression, false information (EIO) and social media platforms have relatively few special rules. As explained above, the legal tools at disposal are commonly used in non-online cases. Furthermore, *liability only applies to users* and those who create their content themselves *but not to social media platforms or internet service providers*, except the provisions of the e-Service Act. A debate arose around the aforementioned court case *Stacho vs. Klub Strážov*,[63] where the plaintiff sought an *apology from the operator of the website, the removal of the post, and damages of EUR 5,000*, and the court in the second instance did not approve damages for the plaintiff but ordered the operator of the website to remove an unlawful comment. In the first instance, the court decided to remove the comment and the compensation of damages for the plaintiff. The question is whether the website operator is responsible for the damages if it had not removed the comment based on the request of the plaintiff. The plaintiff argued that the operator should have acted solely on his request, as the unlawfulness of the comment was clear in his opinion, while the operator argued that he was not aware of the unlawfulness of the comment until the decision of the court in the first

62 Available at: https://bit.ly/3ljd57o.
63 See: Husovec, 2012.

instance. The dispute has yet to be resolved, as the case is at the Supreme Court with no final decision.

In accordance with this, three levels of liability can generally be defined in the Slovak legal system: civil, criminal, and administrative. As stated above, Slovak regulations consider the disputes between the user vs. user and user vs. provider to have basis in private (civil) law, in which legal disagreements are resolved by courts—if a dispute shows elements of criminal or administrative unlawfulness, then the organs of criminal investigation and/or administrative organs can be involved. Therefore, civil liability in this regard is governed by the same rules and regulations as offline cases. Furthermore, there are no known court cases concerning the removal of user posts (or banning users) by social networks.

6.3.1. Rules of criminal liability

The Act no. 300/2005 on Criminal Code (*zákon č. 300/2005 Z. z. Trestný zákon*) or in short: Criminal Code[64] enumerates several provisions that could be applied to natural persons for deeds, which were conducted on online forums. These criminal rules naturally represent restrictions against freedom of expression but are in accordance with constitutional provisions. Crimes that can be committed in online spaces are usually tied to the phenomenon of "hate speech," although this term is never used in legal sources in Slovakia. The crimes connected to users in social network activities are as follows: *a)* disseminating false news[65]; *b)* defamation of nation, race, and belief[66]; *c)* incitement of national, racial, and ethnic hatred[67]; *d)* violence against a group of citizens and against an individual[68]; *e)* supporting and promoting groups aimed at suppression of fundamental rights and freedoms[69]; *f)* manufacturing, possession, and dissemination of extremist materials[70]; *g)* defamation[71]; *h)* unauthorized use of personal data[72]; *i)* serious threats[73]; *j)* dangerous persecution[74]; *k)* harm done to the rights of another[75]; *l)* breach of confidentiality of spoken utterance and other personal expression[76]; *m)* condoning a criminal offense.[77] From this perspective, the following crimes are particularly relevant:

64 Ďuračová, 2005.
65 Section 361 of the Criminal Code.
66 Section 423 of the Criminal Code.
67 Section 424 of the Criminal Code.
68 Section 359 of the Criminal Code.
69 Section 421 of the Criminal Code.
70 Section 422a-422c of the Criminal Code.
71 Section 373 of the Criminal Code.
72 Section 374 of the Criminal Code.
73 Section 360 of the Criminal Code.
74 Section § 360a of the Criminal Code.
75 Section 375-376 of the Criminal Code.
76 Section 377 of the Criminal Code.
77 Section 338 of the Criminal Code.

a) *Disseminating false news* is committed by any person who deliberately creates serious concerns among the population of a certain location or at least a part thereof by disseminating false or alarming news or committing other similar acts capable of giving rise to such danger. The offender shall be liable to a term of imprisonment of up to two years. Any person who reports false or alarming news, or other similar acts referred to above, to a legal entity, the police force, another state authority, or the mass media, although they know that such news is false and may cause serious concerns among the population of a certain location or at least a part thereof, shall be liable to a term of imprisonment of one to five years. Furthermore, any person who, in a crisis situation—even through negligence—creates the danger of serious concern, a mood of despondency, or defeatism among at least a part of the population of a certain location by spreading false or alarming news, shall be liable to a term of imprisonment of between six months and three years.
b) *Defamation of nation, race, and belief* refers to public defamation of any nation, its language, any race or ethnic group, or any individual or a group of persons because of their affiliation to any race, nation, nationality, complexion, ethnic group, family origin, religion or because they have no religion. In this case, the sentence shall be a term of imprisonment of one to three years. The offender shall be liable to a term of imprisonment of two to five years if they commit the offense with at least two more persons, in association with a foreign power or foreign agent, in the capacity of a public official, under a crisis situation, or with a specific motivation.
c) *Incitement of national, racial, and ethnic hatred*: any person who publicly threatens an individual or a group of persons because of their affiliation to any race, nation, nationality, complexion, ethnic group, family origin, or religion, if they constitute a pretext for threatening on the aforementioned grounds, by committing a felony, restricting their rights and freedoms, or who made such restrictions, or who incite the restriction of rights and freedoms of any nation, nationality, race, or ethnic group, shall be liable to a term of imprisonment of up to three years. The same sentence shall be imposed on any person who associates or assembles with others with a view to committing the offense, which shall be liable to a term of imprisonment of two to six years if they commit the offense referred to in association with a foreign power or foreign agent, in public, with a specific motivation, in the capacity of a public official, in the capacity of a member of an extremist group, or in a crisis situation.
d) The crime of *incitement, defamation, and threatening persons because of their affiliation with a race, nation, nationality, complexion, ethnic group, or family origin* is constituted when a person publicly incites to violence or hatred against a group of persons or an individual because of their affiliation to any race, nation, nationality, complexion, ethnic group, family origin, or

their religion, if they constitute a pretext for the incitement on the aforementioned grounds, or defames such group or individual, or threatens them by exonerating an offence deemed to be genocide, a crime against humanity or a war crime, or an offence deemed to be a crime against peace, a war crime, or a crime against humanity, if such crime was committed against such group of persons or individual, or if a perpetrator of or abettor to such crime was convicted by a final and conclusive judgement rendered by an international court, unless it was made null and void in lawful proceedings, publicly denies or grossly derogates such offence, if it has been committed against such person or individual, shall be liable to a term of imprisonment of one to three years.

e) *Violence against a group of citizens or against an individual:* Any person who threatens a group of citizens with killing, inflicting grievous bodily harm, or other aggravated harm, or with causing large-scale damage, or who uses violence against a group of citizens, shall be liable to a term of imprisonment of up to two years. The offender shall be liable to a term of imprisonment of between six months and three years if they commit the offense with a specific motivation, in a more serious manner, or in public.
f) *Supporting and promoting groups aimed at suppressing fundamental rights and freedoms*: Any person who supports or makes propaganda for a group of persons or movements that, using violence, the threat of violence, or the threat of other serious harm, demonstrably aims to suppress citizens' fundamental rights and freedoms shall be liable to a term of imprisonment of one to five years. The offender shall be liable to a term of imprisonment of four to eight years if they commit the offense in public, in the capacity of a member of an extremist group, acting in a more serious manner, or in a crisis situation.
g) *Manufacturing, possession, and dissemination of extremist materials*: Any person who manufactures or disseminates extremist materials or participates in such manufacturing shall be liable to a term of imprisonment of three to six years. The offender shall be liable to a term of imprisonment of four to eight years if they commit the offense acting in a more serious manner, in public, or in the capacity of a member of an extremist group. In the case of possessing extremist material, the offender shall be liable to a term of imprisonment of up to two years.
h) *Defamation*: Any person who communicates false information about another likely to considerably damage the respect of fellow citizens for such a person, damage their career and business, disturb their family relations, or cause serious harm, shall be liable to a term of imprisonment of up to two years. The offender shall be liable to a term of imprisonment of one to five years if they commit this offense and cause substantial damage, by reason of specific motivation, in public or in business acting in a more serious manner. The offender shall be liable to a term of imprisonment of three to eight years if they commit

the offense and cause large-scale damage or causes another to lose their job, collapse their undertaking, or divorce their marriage.

i) *Unauthorized use of personal data*: Any person who, without lawful authority, communicates, makes accessible, or discloses personal data of another obtained in connection with the execution of public administration or with the exercise of constitutional rights of a citizen, or personal data of another obtained in connection with the execution of their own profession, employment, or function, and thus breaches their own obligation prescribed by a generally binding legal regulation, shall be liable to a term of imprisonment of up to one year. The offender shall be liable to a term of imprisonment of up to two years if they commit the offense and causes serious prejudice to the rights of the person concerned, in public, or in a more serious manner.

j) *Serious threats:* Any person who threatens another with killing, inflicting grievous bodily harm, or other aggravated harm to an extent that may give rise to justifiable fears shall be liable to a term of imprisonment of up to one year. The offender shall be liable to a term of imprisonment of between six months and three years if they commit the offense in a more serious manner, against a protected person, with the intention of preventing or obstructing the exercise of fundamental rights and freedoms by another, by reason of specific motivation, or in public.

6.3.2. Rules of administrative liability

As no legislation explicitly and exclusively regulates issues related to social media, there is also no unified system of administrative offenses committed online. However, two branches of administrative law have a closer connection to social media platforms or Internet regulation in general.

The first is *personal data protection*, which has a constitutional basis, as the Constitution in Art. 19 (3) states in general that "*everyone has the right to protection against unauthorized collection, disclosure and other misuse of his or her personal data*." The Slovak Act on Personal Data Protection defines personal data in an almost identical wording to that of Art. 4 (1) of the GDPR.[78] It can be stated that there is no substantive difference between the two pieces of legislation regarding the definition of personal data. Slovak regulations do not explicitly contain data management rules for social media providers. Thus, these are also subject to the general rules set out in the GDPR, the content of which has been taken over by the Slovak legal system and virtually unchanged. Pursuant to Section 110 (1) of the Act on Personal Data Protection, *the Office for Personal Data Protection* acts as the

78 Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

supervisory authority in the field of data protection. Apart from the courts, there is no other body that can play a meaningful role in data protection in relation to social media.

Slovak courts have dealt with violations in which personal data has become illegally available on the Internet. On several occasions, local governments or other persons in the role of data controllers have been fined by the data protection office, after which the case has been brought to court. An example is decision 1S/243/2017 when the District Court in Bratislava upheld a fine imposed on a municipality by a data protection office because it illegally displayed personal data about affected persons on its own website. In another decision – 6S/96/2019 District Court, the Court annulled a decision of the Office for Personal Data Protection imposing an unusually high penalty (€ 25,000) on an individual who illegally disclosed personal data as an operator of a publicly available directory, which also occurred on its own website. In this case, the court did not refute the infringement itself, but found the fine imposed by the data protection office to be too severe. *Based on the available sources, no court decision specifically addresses data breaches on social media platforms or search engine providers.*[79]

The Office for Personal Data Protection has not issued its own guidelines or other documents addressing the protection of personal data in relation to social media. Only Guidelines 8/2020 on the targeting of social media users,[80] developed by the European Data Protection Board, can be found on the office's website. According to the available sources, the Slovak data protection authority has addressed at least one case of a data breach related to the services of community platforms. In the known case, the data controller (institution providing childcare services) published a photograph of a child on a community platform with the prior consent of the child's legal representative. The legal representative later requested the deletion of the photograph from the social media interface, which was rejected by the data controller, who later argued that the photograph was needed for criminal proceedings. However, in the opinion of the Office for Personal Data Protection, the child's right to data protection in the given case took precedence over the legitimate interests of the data controller. Thus, this constituted a violation of Art. 5 (1) (a) GDPR, adding that the legal basis of the data was subject to display.[81] Furthermore, *if illegal content appears on a website that violates the data protection rules, the Office may oblige the data controller to take* measures to remedy the identified deficiencies (in this case, to delete the illegal content).[82]

Aside from personal data protection, electoral rules contain the soma aspects of political campaigns on social networks. The election campaigns and the order of elections are regulated by Act no.181/2014 on the election campaign. (*zákon č.*

79 See: https://bit.ly/3tBGkWF.
80 See: https://bit.ly/392k7HN.
81 See: https://bit.ly/3Aam9S3.
82 Pursuant to Section 99 of the Act on Personal Data Protection.

181/2014 Zb. o volebnej kampani)[83] and Act no. 180/2014 on the conditions for exercising the right to vote. (*zákon č. 180/2014 Zb. o podmienkach výkonu volebného práva*)[84]. The Ministry of the Interior is responsible for conducting elections; some oversight functions are performed by *the State Committee for the Supervision of Electoral and Political Party Financing*, but ultimately by the Supreme Administrative Court[85] and the Constitutional Court.[86] The Slovak legislation is based on the concept of a closed election campaign,[87]—an election campaign can only be conducted by certain legal entities—so above all by the political parties and their representatives participating in the election; third parties may only participate in an election campaign with prior registration. The active participation of other people in the election campaign is prohibited.

The basic requirement for political advertising during an election campaign is that it is transparent, meaning that the voter can clearly identify the nature of political advertising, which political party has created the advertisement, and the identity of the advertising agency. This restriction also applies to opinion polls.[88] An election campaign can only be conducted during the election campaign period, which lasts until 48 hours before the election announcement, after which there is campaign silence.[89] In addition, political advertising on radio and television may only be broadcast during the period set aside for that purpose—between 21 and 48 hours before the elections[90]—and the results of election polls may no longer be made public from the 14th day before the elections.[91] The law also regulates the timeframes for election advertisements on radio and television. In this context, *freedom of expression can be exercised within significant limits in the context of political advertising.*[92] It is therefore interesting that *the law explicitly excludes its own applicability to online media*, so under § 12 (6) and § 14 (2) of the Act on Election Campaign, the political campaigns restrictions – with the exception of the rules on

83 This legislation contains detailed regulations on election campaigns, such as who can conduct election campaigns and political agitation under what conditions. See: https://bit.ly/394fDAu.

84 This law contains the rules for the conduct of elections, how the right to vote can be exercised, and what operational tasks each state body has in the conduct of elections. See: https://bit.ly/3hvmnMl.

85 Pursuant to Art. 142 (2) of the Constitution, the Supreme Administrative Court decides on the legality and constitutionality of local elections.

86 Pursuant to Art. 129 (2) of the Constitution, the Constitutional Court decides on the legality and constitutionality of the presidential, parliamentary, and European elections.

87 Orosz, 2016, pp. 105–106.

88 Section 15 of the Act on Election Campaign: Everyone who is running an election campaign is obliged to ensure that political advertisements, paid advertisements, published election posters, and all other ways of conducting an election campaign contain information about the customer and producer; the same applies to present pre-election and opinion polls.

89 Section 2(2) of the Act on Election Campaign: The election campaign begins on the day of the publication of the decision to declare the election in the Collection of Laws of the Slovak Republic and ends 48 hours before the day of the election.

90 Section 12 of the Act on Election Campaign.

91 Section 17 of the Act on Election Campaign.

92 There has been a wider academic debate around these limitations. See Orosz, 2016, pp. 105–106.

transparency – do not apply to "transmissions over the internet." Regarding the requirements of the rules and restrictions *on the publication of the results of opinion polls, the law does not contain any provision for internet media*. Restrictions on a certain level of freedom of expression during the election campaign do not concern substantive issues (the Committee or the ministry does not check who said what during the election campaign) but are aimed at complying with formal conditions (e.g., registration obligation, monitoring, breaches of campaign silence, etc.). The Committee may impose a fine of between €1,000 and €300,000 to a political party or a candidate that breaches the campaign silence or discloses the results of a poll.[93] Ultimately, however, the Committee cannot interfere in political communication, so it cannot judge whether a message contained in political advertising violates constitutional and legal restrictions on freedom of expression, provided that political advertising is formally lawful. *The state committee for the supervision of electoral and political party financing has no power to sanction the unauthorized deletion of content on social or other media.*

Further discussions must concern the field of misdemeanors, such as misdemeanors against civil society,[94] among which are offenses committed by a person who *a)* injures the honor of another by insulting or ridiculing him; *b)* intentionally makes a false or incomplete statement to a public authority, a municipal authority, or an organization for the purpose of obtaining an unjustified advantage, and *c)* intentionally disrupts civil coexistence by threatening bodily harm, minor bodily harm, false accusations of misconduct, endorsements, or other abusive behavior. Such unlawful behavior can be fined up to €331. Other types of non-criminal offenses in the online space are represented by misdemeanors of extremism,[95] which can be committed when a person: *a)* uses in public a written, graphic, pictorial, visual, audio, or audio-visual representation of texts and statements, flags, badges, slogans or symbols of groups or movements and their programs or ideologies that are directed towards the suppression of fundamental human rights and freedoms; *b)* uses in public written, graphic, pictorial, visual, audio or visual-sound design advocating, supporting or inciting hatred, violence, or unjustifiably different treatment against a group of persons or an individual because of their membership of a race, nation, nationality, color, ethnic group, descent, or religion. This behavior can be fined up to €500.

Considering the individual administrative branches, unwanted advertisements must also be taken into account. *The freedom of expression includes the right to disseminate information, which has a commercial character* in the interest of the promotion of certain products, which includes the dissemination of such information via the Internet.[96] This is regulated by the provisions of the *e-Services Act* and falls into

93 That is, 48 hours before the election. See: Section 2(2) of the Act on Election Campaign.
94 Section 49 of the Act on Misdemeanors.
95 Section 47a of the Act on Misdemeanors.
96 Jakab, 2016, pp. 171–172.

the category of *unsolicited commercial communications* (which are the main regulations of this act). This is a negative phenomenon for several reasons, mainly due to threats to privacy, customer fraud, and risk to minors and adolescents. In addition, a significant portion of spam has a deceptive or even fraudulent nature, contains pornographic material, unreasonable violence, or incitement to hatred.[97] The problematic act itself, in some parts, falls under the GDPR regulation, but the e-Services Act together with the e-Commerce Directive provide a relatively complex regulation, although this may be subject to change with the planned *Digital Services Act* and *Digital Markets Act.*[98]

7. Closing remarks

Freedom of expression is one of the constitutional cornerstones of a democratic society. The social and technological developments of the last decade made it clear that *the state may not continue to take a passive attitude towards the freedom of speech*, as this is not sufficient to ensure only that the state itself does not intervene in the exchange of information of citizens. Instead, *it must actively guarantee and ensure the realization of freedom of expression and exchange of information*.

Social media is unregulated in Slovakia, and there is currently no legislative intention to regulate it. The scope of the current regulations covers the provision of information society services. Regarding the responsibility of the service provider for content control, the Slovak legislation transposes Art. 14 of the e-Commerce Directive with practically no substantive changes. Thus, under Slovak law, a service provider can be held liable if it has not removed such content after becoming aware of the infringement unless it has produced the content itself or has a significant influence on its production. The service provider has no obligation to monitor the content, and the regulation explicitly prohibits the service provider from searching users' data.

There are no regulations of *the alleged or real censorship of social media platforms*: the main legal doctrine and the Constitutional Court do not define censorship as a phenomenon that can occur between two private entities. *Censorship is only considered an action from the state against freedom of expression,* which some authors consider to be outdated. In this manner, *de facto censorship* differs from *de iure censorship*, which is a narrower term. As a rule, an individual can go to court in the event of the removal of infringing content and an infringement suffered online. In some sectors, such as those concerning the protection of personal data and copyright,

97 Jakab, 2016, pp. 173–174.
98 See: The Digital Services Act package [Online]. Available at: https://bit.ly/3AkREJ6.

there is an administrative supervisory body, including administrative intervention under sectoral legislation.

In the category of "Fake News," the official viewpoint of the Slovak government is that the state must strengthen its own means and capacities for resilience to information operations and cooperate with experts from the public and private sectors to detect and analyze false information. Based on this, a government plan was created (but not yet implemented) to strengthen state reactions to various elements of information operations, reacting primarily to false reports, hoaxes, conspiracy theories, disinformation, and malinformation. Whether such operations can be carried out effectively in accordance with human rights legislation, freedom of speech, unrestricted access to information, and basic human freedoms guaranteed by the Constitution of the Slovak Republic is yet to be seen.

Bibliography

BARTOŇ, M. (2002) *Svoboda projevu a její meze v právu České republiky***,** Praha: Linde, a.s.

Delfi AS v. Estonia, [Online]. Available at: https://bit.ly/3A8wNsL (Accessed: 31 May 2021).

DRGONEC, J. (2015) 'Zákaz cenzúry podľa Ústavy Slovenskej republiky: Implikované základné právo alebo ústavný princíp a súvisiace otázky' *Právník*, 1., pp. 61-79. [Online]. Available at: https://bit.ly/3C2tuUi (Accessed: 31 May 2021).

European Democracy Action Plan [Online]. Available at: https://bit.ly/39iOMRv (Accessed: 31 May 2021).

ĎURAČOVÁ, M. (2005) Translation of the Slovak Criminal Code [Online]. Available at: https://bit.ly/3z79dv8 (Accessed: 31 May 2021).

Európsky parlament žiada jasnú legislatívu pri odstraňovaní obsahu na sociálnych sieťach [Online]. Available at: https://bit.ly/3lnG1uH (Accessed: 31 May 2021).

FILIP, J. (1998) 'Dogmatika svobody projevu z hlediska teorie, legislativy a soudní praxe' *Časopis pro právní vědu a praxi*, Vol. 4, pp. 618–637.

Final results of the Eurobarometer on fake news and online disinformation [Online]. Available at: https://bit.ly/3C7mrtv (Accessed: 31 May 2021).

Guidelines 8/2020 on the targeting of social media users [Online]. Available at: https://bit.ly/3A4c9d4 (Accessed: 31 May 2021).

HUSOVEC, M. (2012) *KS Trenčín: Zodpovednosť poskytovateľa diskusného fóra za údajne difamačné príspevky tretích osob*, [Online]. Available at: https://www.lexforum.sk/405 (Accessed: 31 May 2021).

Information Disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe report DGI(2017)09 [Online]. Available at: https://bit.ly/2X8DSuB (Accessed: 31 May 2021).

KLEIN, T. (2018) 'Az online diskurzusok egyes szabályozási kérdései' in: Klein, T. (eds.) *Tanulmányok a technológia- és cyberjog néhány aktuális kérdéséről*. Budapest: Médiatudományi Intézet. [Online]. Available at: https://nmhh.hu/dokumentum/194190/MK30web.pdf (Accessed: 31 May 2021).

KOLTAY, A. (2019) 'A social media platformok jogi státusa a szólásszabadság nézőpontjából'**,** *In Medias Res.* [Online]. Available at: https://bit.ly/3zazR6d (Accessed: 31 May 2021).

Koľko Slovákov je na sociálnych sieťach? [Online]. Available at: https://bit.ly/3z79wpM (Accessed: 31 May 2021).

Králi sociálnych sietí na Slovensku: Facebook, YouTube a Instagram [Online]. Available at: https://bit.ly/3EobOEO (Accessed: 31 May 2021).

MAJERČÁK, T. (2016) 'Obmedzenie slobody prejavu a práva na informácie v zákone o volebnej kampani' in Majerčák, T. (eds.*) Sloboda prejavu a jej limity – IV. ústavné dni*. Košice: UPJŠ [Online]. Available at: https://www.upjs.sk/public/media/17625/05_SlobodaPrejavu.pdf (Accessed: 31 May 2021).

MORAVEC, O. (2013) *Mediální právo v informační společnosti*, Praha: Leges.

PAVLÍČEK, V. ET AL. (1999) *Ústava a ústavní řád. 2. Díl,* Praha: Linde, a.s.

Nález Ústavného súdu Slovenskej republiky sp. zn. II. ÚS 307/2014 z 18. decembra 2014 [Online]. Available at: https://bit.ly/3C8U2n6 (Accessed: 31 May 2021).

RADOMÍR, J. 'SLOBODA PREJAVU A SPAM' IN MAJERČÁK, T. (eds.*) Sloboda prejavu a jej limity – IV. ústavné dni*. Košice: UPJŠ [Online]. Available at: https://bit.ly/3ntTgwL (Accessed: 31 May 2021).

Orosz, L. (2016) 'Sloboda prejavu vo volebnej kampani' in Majerčák, T. (eds.*) Sloboda prejavu a jej limity – IV. ústavné dni*. Košice: UPJŠ [Online]. Available at: https://www.upjs.sk/public/media/17625/05_SlobodaPrejavu.pdf, (Accessed: 31 May 2021).

Podľa Klusa by mali sociálne siete prijať väčšiu zodpovednosť: Musíme bojovať proti dezinformáciám [Online]. Available at: https://bit.ly/3nv9F3R (Accessed: 31 May 2021).

Programové vyhlásenie Vlády SR [Online]. Available at: https://rokovania.gov.sk/RVL/Material/24756/1 (Accessed: 31 May 2021).

Sociálne siete musia na seba prebrať viac zodpovednosti, zhodli sa Bilčík a Šimečka [Online]. Available at: https://bit.ly/3z8S58e (Accessed: 31 May 2021).

Správa o stave ochrany osobných údajov za obdobie 25. MÁJ 2018 až 24. MÁJ 2019 [Online]. Available at: https://bit.ly/3C6lMZf (Accessed: 31 May 2021).

Správa o stave osobných údajov za rok 2020, Úrad na ochranu osobných údajov Slovenskej republiky [Online]. Available at: https://bit.ly/3tGV1aT (Accessed: 31 May 2021).

Vládny materiál: Koordinovaný mechanizmus odolnosti Slovenskej republiky voči informačným operáciám [Government material: Coordinated mechanism of the Slovak Republic's resilience to information operations] [Online]. Available at: https://bit.ly/3A8TcpL.

Chapter IX

Freedom of Expression on Social Networks: An International Perspective

Dušan V. Popović

1. Legal aspects of content censorship on social networks

Social networks are omnipresent; yet, there is no generally accepted definition of them. In order to define 'social networks' for our current purposes, we have identified several common features of existing social media platforms, which are presented in the literature.[1] First, social networks are Web 2.0-based applications. The shift to Web 2.0 applications can be described as a shift from the user as a consumer to the user as a participant. These apps are designed to enable users to interact, create, and share content online. Second, user-generated content is the essential (but not exclusive) component of social networks. The notion of 'user generated content' is not limited to text, photos, or videos; it could well be a simple 'like.' Third, social networks connect user-specific profiles with those of other individuals or groups. User profiles are thus the pillars of every social network. The manner in which users identify themselves may vary, but every social network tracks users' Internet Protocol (IP) address. Given their similarities from a freedom of speech perspective, we shall take the same approach *stricto sensu* to social networks, such as Facebook or Twitter, and video-sharing portals, such as YouTube.

Analyzing the legal aspects of content censorship on social networks starts with the examination of the foundations of freedom of speech (Section 1.1), as well as the very

1 See for example: Obar and Wildman, 2015, pp. 745–750.

Dušan V. Popović (2021) Freedom of Expression on Social Networks: An International Perspective. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 277–310. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_9

notion of 'speech,' which is extensively interpreted in both an offline and online context (Section 1.2). In the first years following their creation, social networks have legally been considered private spaces. The next section examines whether they should be considered as public forums, given their social function (Section 1.3). The paper will also examine the legal basis for content censorship in comparative law. There are two main approaches to the regulation of social networks, which serve as models for other jurisdictions: the US and the EU models (Section 1.4). Further to government regulation of social networks, we witness different forms of internal rules and regulations adopted by social networks, such as terms of service, privacy policies, IP policies, and community standards (Section 1.5). However, there are two main downsides of such self-regulation: the loss of equal access to speech and the lack of accountability (Section 1.6).

1.1. The foundations of freedom of speech on the Internet

Freedom of speech allows ordinary people to participate in the spread of ideas. It undoubtedly represents an important element of democratic culture, in the sense that everyone, not only the political or cultural elite, has a chance to participate in public dialogue. Freedom of speech is interactive, since exposure to someone else's ideas influences and potentially reshapes us. Freedom of speech is also appropriative in the sense that every participant relies on, draws ideas from, and modifies and/or criticizes the existing cultural background.

The theoretical foundations of freedom of speech can be categorized in different ways.[2] Freedom of speech may be understood as a means of truth discovery. According to John Stuart Mill, the recognition of truth is a prerequisite of social development. Therefore, the limitation of freedom of speech is inadmissible, since the restricted opinion may carry the truth.[3] On the other hand, freedom of speech can be seen as an instrument of democratic self-government. According to Alexander Meiklejohn and many others, freedom of speech enables the proper operation of society. Another line of thought sees freedom of speech as a value in itself—a right to which every citizen is entitled. Ronald Dworkin is a notable representative of this individualist theory.

These theories are reflected in the case law of national and supranational courts. In the United States, the US Supreme Court adopted a landmark decision in 1964 in the case *New York Times Co. v. Sullivan*, restricting public officials' ability to sue for defamation.[4] Specifically, the court held that if a plaintiff in a defamation lawsuit is a public official or a person running for public office, not only must they prove the normal elements of defamation, i.e., publication of a false defamatory statement to a third party, they must also prove that the statement was made with actual malice, meaning that the defendant either knew the statement was false or recklessly disregarded its

2 For a more detailed analysis of different theoretical justifications of freedom of speech, see: Koltay, 2019, pp. 8–15.

3 John Stuart Mill laid down the foundations of freedom of speech in his essay *On Liberty* (1859).

4 US Supreme Court, *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964).

veracity. On the other side of the Atlantic, European (national) courts' case law is under the significant influence of the views and interpretations expressed by the European Court of Human Rights (ECtHR). The right to freedom of expression, guaranteed under Article 10 of the European Convention on Human Rights,[5] is interpreted to include the right to freely express opinions, views, and ideas, and seek, receive, and impart information regardless of frontiers. Freedom of expression is applicable not only to information or ideas that are favorably received or regarded as inoffensive, but also to those that may offend or disturb. In its landmark decision in *Handyside v. the United Kingdom,* the ECtHR defined freedom of expression as one of the essential foundations of a democratic society and a basic condition for its progress and for the development of every man.[6] As noted in the Council of Europe's Guide to Human Rights for Internet Users[7] and its explanatory memorandum, the ECtHR has affirmed in its jurisprudence that Article 10 is fully applicable to the Internet.[8] Member states have a primary duty, pursuant to Article 10 ECHR, not to interfere with the communication of information between individuals, be they legal or natural persons.

The global expansion of the Internet has provided a means by which free speech can reach broader audiences than ever before. The Internet's technological superiority and affordability facilitate citizens' participation in information exchange. However, the majority of Internet users exercise their right to freedom of expression anonymously, which can lead to certain abuses or even criminal offenses that could *de facto* be impossible to persecute.

1.2. The concept of speech

International and national legal documents do not use uniform terminology to designate the right to participate in public debate. The First Amendment of the United States Constitution, adopted in 1791, employs the term 'freedom of speech:'

> Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.

It has been heavily debated whether the free speech and free press clauses are coextensive or whether one reaches where the other does not. Justice Stewart argued that the fact that the First Amendment speaks separately of freedom of speech and

5 Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, 1950.

6 ECtHR, *Handyside v. the United Kingdom,* 7 December 1976, § 49.

7 Council of Europe, Recommendation of the Committee of Ministers to Member States on a Guide to Human Rights for Internet users, CM/Rec(2014)6, 16 April 2014.

8 See for example: ECtHR, *Perrin v. the United Kingdom*, 18 October 2005; ECtHR, *Renaud v. France*, 25 February 2010; ECtHR, *Editorial Board of Pravoye Delo and Shtekel v. Ukraine*, 5 May 2011.

freedom of the press is no accident, but an acknowledgment of the critical role the press plays in US society. In his view, the Constitution requires sensitivity to that role and to the press's special needs in performing it effectively.[9] However, contemporary interpretations of the First Amendment analyze the speech and press clauses under an umbrella 'freedom of expression' standard. The French Declaration of the Rights of Man and of the Citizen (*Déclaration des droits de l'homme et du citoyen*), adopted in 1789, employs the term 'freedom to express thoughts and opinions:'

> The free communication of thoughts and opinions is one of the most precious of the rights of man. Every citizen may, accordingly, speak, write, and print with freedom, but shall be responsible for such abuses of this freedom as shall be defined by law.

More recently adopted legal documents employ the term 'freedom of expression' rather than 'freedom of speech.' For example, the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR), adopted in 1948 and 1966 respectively, both state that individuals have a right to freedom of expression; this right includes the freedom to seek, receive, and impart information and ideas of all kinds.[10] The European Convention on Human Rights also employs the term 'freedom of expression.'[11]

The concept of 'freedom of speech' has been interpreted extensively, so as to include not only direct speech (words) but also symbolic speech (actions). In the United States, the freedom of speech includes *inter alia* the right not to speak,[12] the right to use certain offensive words and phrases to convey political messages,[13] the right to advertise commercial products and professional services,[14] and the right to burn the flag in protest.[15] The ECtHR also considers 'freedom of expression' to cover both direct and symbolic speech. For instance, the Court found that freedom of expression includes artistic expression such as a painting,[16] the production of a play,[17] and information of a commercial nature.[18] With regard to the so-called 'negative right' not to express oneself, the ECtHR does not rule out that such a right is protected under the European Convention on Human Rights, but it has found that this issue should be addressed on a case-by-case basis.[19] Specifically in the context of the Internet, the ECtHR has emphasized that Art. 10 of the Convention is to apply to communication

9 *Houchins v. KQED*, 438 U.S. 1, 17 (1978) (concurring opinion).
10 UDHR, art. 19; ICCPR, art. 19.
11 ECHR, art. 10.
12 *West Virginia Board of Education v. Barnette*, 319 U.S. 624 (1943).
13 *Cohen v. California,* 403 U.S. 15 (1971).
14 *Bates v. State Bar of Arizona*, 433 U.S. 350 (1977).
15 *Texas v. Johnson*, 491 U.S. 397 (1989); *United States v. Eichman*, 496 U.S. 310 (1990).
16 ECtHR, *Müller and Others v. Switzerland*, 24 May 1988.
17 ECtHR, *Ulusoy and Others v. Turkey,* 25 June 2019.
18 ECtHR, *Casado Coca v. Spain,* 24 February 1994.
19 ECtHR Guide, 2020, p. 14.

on the Internet, whatever the type of message being conveyed and even when the purpose is profit making in nature.[20]

The Internet has undoubtedly introduced new forms of communication, i.e., new forms of opinion expression. For example, a 'like' on a social network is a form of speech, as it represents an Internet user's statement. This was established in the case *Bland v. Roberts*, where a public sector employee sued because he was fired for clicking the Facebook 'like' button on his employer's re-election rival's campaign website. The judge dismissed the free speech claim stating that 'liking' web content is not 'sufficient' speech to warrant constitutional protection. However, the Fourth Circuit reversed the decision on the First Amendment issue, holding that:

> On the most basic level, clicking on the 'like' button literally causes to be published the statement that the user 'likes' something, which is itself a substantive statement. In the context of a political campaign's Facebook page, the meaning that the user approves of the candidacy whose page is being liked is unmistakable. That a user may use a single mouse click to produce that message that he likes the page instead of typing the same message with several individual key strokes is of no constitutional significance.[21]

The court also noted that the act of 'liking' a page itself results in an affirmative statement made by a Facebook user to their friends. Consequently, choosing to 'like' something on Facebook produces speech.[22]

The US courts also held that the First Amendment protects as 'speech' the results produced by an Internet search engine. In *Search King, Inc. v. Google Technology, Inc.*, the court concluded that Google's page rankings were subjective results that constituted 'constitutionally protected opinions' entitled to full constitutional protection.[23] Likewise, in *Langdon v. Google, Inc.*, the court refused to order Google and Microsoft to prominently list the plaintiff's site in their search results, reasoning that:

> The First Amendment guarantees an individual the right to free speech, 'a term necessarily comprising the decision of both what to say and what not to say.' (…) The injunctive relief sought by plaintiff contravenes defendants' First Amendment rights.[24]

Just as newspapers cannot be forced to print editorial content or advertising, the court held that search engines cannot be forced to include links that they wish to exclude. This full protection remains when the choices about how to select and arrange the material are implemented with the help of computerized algorithms.[25]

20 ECtHR, *Ashby Donald and Others v. France,* 10 January 2013.
21 *Bland v. Roberts*, No. 12-1671, 4th Cir., 18 September 2013.
22 For an extensive analysis of *Bland v. Roberts* case see: Sarapin and Morris, 2014, pp. 131-157.
23 No. CIV-02-1457-M, 2003 WL 21464568, at *4, W.D. Okla. 27 May 2003.
24 474 F. Supp. 2d 622, 629–30 (D. Del. 2007) (citing *Riley v. National Fed'n of the Blind of N.C., Inc.*, 487 U.S. 781, 796–97 (1988); *Miami Herald Pub'g Co. v. Tornillo*, 418 U.S. 241, 256 (1974).
25 Volokh and Falk, 2012, pp. 886–887.

The US legal system differentiates among several categories of speech, some of which do not fall under the freedom of speech protection. The following categories of speech are given lesser or no protection by the First Amendment: obscenity, fighting words, defamation (including libel and slander), child pornography, perjury, blackmail, incitement to imminent lawless action, true threats, solicitations to commit crimes, and plagiarism of copyrighted material. Contrary to the US legal system, the European (national) legal systems and the European Convention on Human Rights do not introduce categories of speech. Instead, they prescribe different limitations on the freedom of speech, such as protection against defamation or speech interfering with the intimate and private sphere, the maintenance of public order and national security, the protection of consumers against misleading commercial messages, the protection of children against materials that are harmful to their development, and the protection of certain social groups against hatred.[26]

1.3. Social networks as a public forum?

Since their inception, social networks such as Facebook and Twitter have been legally considered as private spaces. However, in recent years, social networks are increasingly being perceived as forums of public communication. In line with this tendency, the US courts examined whether the public forum doctrine could be applied to social networks. The nuances of the public forum doctrine were articulated in the case *Perry Education Association v. Perry Local Educators' Association* in 1983.[27] Justice Byron R. White explained three categories of government property for the purposes of access for expressive activities: (1) traditional or quintessential public forums, (2) limited or designated public forums, and (3) non-public forums. According to the public forum doctrine, the government can impose reasonable time, place, and manner restrictions on speech in all three property categories but has limited ability to impose content-based restrictions on traditional or designated public forums.

Nowadays, many politicians choose to set up official Facebook, Twitter, and Instagram accounts to communicate with citizens. These accounts are used for official purposes. Should these social network accounts be perceived as a public forum? In *Knight First Amendment Inst. at Columbia Univ. v. Trump,*[28] a group of seven citizens, represented by the Knight First Amendment Institute, sued US President Trump. Their complaint alleged that when President Trump blocked them on Twitter, he engaged in viewpoint discrimination in a public forum, an action that would violate

26 In certain situations, the ECtHR does not even examine the compatibility of a limitation with Art. 10 of the European Convention on Human Rights. This happens when the ECtHR finds an abuse of the freedom of speech, within the meaning of Art. 17 of the Convention. See: Koltay, 2019, p. 20.

27 460 U.S. 37 (1983).

28 302 F. Supp. 3d 541 (S.D.N.Y. 23 May 2018).

the First Amendment's freedom of speech guarantee. President Trump argued that because this was his private account,[29] created in 2009, it was not subject to First Amendment claims. In 2019, the 2nd and 4th Circuit Courts of Appeals ruled that government use of social media creates a designated public forum, and government officials cannot engage in viewpoint discrimination by blocking comments.[30] The Court found that President Trump violated the First Amendment by removing several individuals who were critical of him and his governmental policies from the 'interactive space' of his Twitter account. The appeals court agreed with the lower court that the interactive space associated with Trump's Twitter account is a designated public forum and that blocking individuals because of their political expression constitutes viewpoint discrimination.[31]

From a freedom of expression perspective, it is particularly relevant to determine whether social networks should be treated as tech or media companies. Social networks, such as Facebook, have repeatedly insisted that their service is a neutral tech platform, not a publisher or a media company. A publisher, after all, could be expected to make factual and qualitative distinctions, and might be responsible, reputationally or legally, for the content it publishes, whereas a platform is nothing but empty space. However, in court proceedings in the United States, when Facebook was sued by an app startup that alleged that Mark Zuckerberg developed a 'malicious and fraudulent scheme' to exploit users' personal data and force rival companies out of business, Facebook's lawyers argued that decisions about what not to publish should be protected because Facebook is a publisher. Facebook's lawyers argued in court that the social network's decisions about data access were a 'quintessential publisher function' and constituted protected activity, adding that this includes both the decision of what to publish and the decision of what not to publish.[32]

If social networks are publishers, then the manner in which they select content results from editorial decisions and should be treated as 'speech.' In addition, if a social network has an opinion, than such an opinion could, under certain legally defined conditions, be restricted.

29 President Trump maintained only one Twitter account that he used for both private and official interactions with American citizens.

30 928 F. 3d 226 – Court of Appeals, 2nd Circuit 2019.

31 The petition for rehearing was denied on 23 March 2020. On 31 July 2020, the Knight Institute filed a second lawsuit in federal court against President Trump and his staff for continuing to block followers from the @realDonaldTrump Twitter account. On 5 April 2021, the Supreme Court vacated the judgment. The case has been remanded to the United States Court of Appeals for the Second Circuit with instructions to dismiss the case as moot, given that Donald Trump is a private citizen now.

32 Sam Levin, 'Is Facebook a publisher? In public it says no, but in court it says yes' *The Guardian* (3 July 2018) at https://www.theguardian.com/technology/2018/jul/02/facebook-mark-zuckerberg-platform-publisher-lawsuit.

1.4. Legal basis for content censorship in comparative law

Typically, liability for third-party content attaches when the disseminator has the discretion to publish it or not. If a disseminator cannot exercise editorial control, the disseminator is not legally responsible for third-party content it had to disseminate. In contrast, if the disseminator can exercise editorial control over the content, the disseminator accepts legal liability for the (editorial) decisions it makes. Online intermediaries, including social networks, do not entirely fit into either category. However, that does not mean that legislators have not imposed certain content-related obligations on them.

We shall analyze two approaches to the regulation of social networks, which serve as models to other jurisdictions: US and EU law. Our comparative analysis shall start with US law, since the United States is the Internet's birthplace. The US model protects intermediaries from liability for distributing third-party user content based on the 'Good Samaritan' rule, with the exception of certain laws: criminal law, intellectual property law, communications privacy law, and sex trafficking law. The US model could be seen as more favorable to online platforms than the EU's approach. The United States' neighboring countries and traditional economic partners follow its approach. For example, the US-Mexico-Canada agreement (USMCA, also known as NAFTA 2.0), concluded in 2018, requires Canada and Mexico to adopt protections in line with US legislation.[33] On the other hand, EU law provides liability exemption in favor of Internet intermediaries, concerning illegal content and activities online. The exemptions from liability only cover cases where the information society service provider's activity is limited to the technical operation process. The EU model is followed not only by EU member states, but also by other European countries that are candidates or potential candidates for EU membership.[34]

1.4.1. US law

The Communications Decency Act of 1996,[35] particularly Section 230, is the most important piece of US legislation related to online speech. The Act is the short name of Title V of the Telecommunications Act of 1996, as specified in Section 501 of the 1996 Act. Title V has affected the Internet and online communications in two significant ways. First, it attempted to regulate both indecency (when available to children) and obscenity in cyberspace. Second, Section 230 of the Communications Act of 1934 (Section 9 of the Communications Decency Act / Section 509 of the

33 Art. 19.17 of the USMCA: "No Party shall adopt or maintain measures that treat a supplier or user of an interactive computer service as an information content provider in determining liability for harms related to information stored, processed, transmitted, distributed, or made available by the service, except to the extent the supplier or user has, in whole or in part, created, or developed the information."

34 See for example: Republic of Serbia, Law on Electronic Commerce, *Official Journal* 41/2009, 95/2013 and 52/2019, Arts. 16–20.

35 47 U.S.C. § 230.

Telecommunications Act of 1996) has been interpreted to mean that operators of Internet services are not traditional publishers. Section 230(c)(1) reads: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." There are three elements to this immunity. First, the immunity applies to a 'provider or user of an interactive computer service.' The courts have interpreted 'providers' extensively to include any service available through the Internet. Furthermore, 'users of interactive computer services' should cover all providers' customers. Second, the immunity applies to any claims that treat the defendant as a 'publisher' or 'speaker.' However, the courts usually interpret this element more extensively so that it applies regardless of whether the claim's *prima facie* elements contain the terms 'publisher' or 'speaker.' Third, immunity applies when the plaintiff's claim is based on information provided by another information content provider, i.e., by a third party.[36]

Section 230 immunity is not unlimited. It has four statutory exclusions where it is categorically unavailable. First, prosecutions of federal crimes (e.g., obscenity, sexual exploitation of children) are not immunized by Section 230. Second, Section 230 does not apply to plaintiffs' claims based on the Electronic Communications Privacy Act (ECPA)[37] or state law equivalents. Third, Section 230 does not apply to claims based on the Fight Online Sex Trafficking Act (FOSTA),[38] related to websites that unlawfully promote and facilitate prostitution and/or facilitate traffickers in advertising the sale of unlawful sex acts involving sex trafficking victims. Fourth, Section 230 does not apply to intellectual property claims. However, the courts differ in interpreting whether this exclusion applies only to federal intellectual property claims or also to state IP claims. In *Perfect 10 v. CCBill*, the Ninth Circuit held that the exclusion only applied to federal intellectual property claims.[39] However, courts outside the Ninth Circuit do not agree with the *CCBill* ruling, so state intellectual property claims are still viable in those jurisdictions.

When discussing the relationship between freedom of speech and IP rules in US law, one should bear in mind that there is also a specific 'notice and takedown' procedure related to copyrighted works, which was introduced by the Digital Millennium Copyright Act (DMCA).[40] This procedure allows a copyright owner to request the removal of content posted online. The DMCA shields online service providers from monetary liability and limits other forms of liability for copyright infringement—referred to as safe harbors—in exchange for cooperating with copyright owners to expeditiously remove infringing content if the online service providers meet certain conditions. Specifically, Subsection 512(c)(1)(A) of the DMCA requires that the service provider: (1)

36 For an overview of US case-law see: Balasubramani, 2016/2017, pp. 275–286.

37 18 U.S.C. §§ 2510–2523. The ECPA was significantly amended by the Communications Assistance to Law Enforcement Act (CALEA) in 1994, the USA PATRIOT Act in 2001, the USA PATRIOT Reauthorization Acts in 2006, and the FISA Amendments Act of 2008.

38 Public Law No: 115-164, 11 April 2018.

39 *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102, 9th Cir. 2007.

40 The DMCA safe harbors, codified at 17 U.S.C. § 512, are part of the Copyright Act.

does not have actual knowledge that the material or an activity using the material on the system or network is infringing; (2) in the absence of such actual knowledge, is not aware of facts or circumstances from which infringing activity is apparent; or (3) upon obtaining such knowledge or awareness, acts expeditiously to remove, or disable access to, the material. The DMCA has become a *de facto* global standard for addressing online copyright infringements, since the vast majority of removal requests are sent to global platforms that are US-based companies subject to the DMCA.

The DMCA offers Internet service providers protection from copyright liability if they expeditiously remove material in response to (essentially unverified) infringement complaints. Even if the accused poster responds with counter-notification of non-infringement,[41] the DMCA requires that the service provider keep the post offline for more than a week. Obviously, this procedure can be abused for censorship purposes. Indeed, the threat of secondary liability induces service providers to comply with the DMCA's notice and takedown provisions, making it more difficult for speakers to post material that challenges someone who can potentially make a copyright claim.[42] Since the notice and takedown procedures are implemented in a non-transparent way,[43] it is difficult to track such abuse. Moreover, because the notice and takedown procedures involve immediate removal but lack any legal oversight, there are no effective means to protect against abuse of the process. As long as the automatic enforcement system does not distinguish legitimate removal requests from non-copyright requests, there is great potential for misuse.[44] However, the DMCA does not impose a general filtering obligation, as the service provider is not required to block an allegedly infringing file from being re-uploaded to its service after the file has been taken down in response to a copyright owner's notice.[45]

1.4.2. EU law

The US DMCA legislation inspired the EU to enact the Directive on Electronic Commerce,[46] including safe harbors for mere conduits, caching, and hosting.[47] The

41 A mechanism that allows a user to contest the removal request.

42 Seltzer, 2010, p. 177.

43 The notice-and-takedown procedure is administered by private companies. Unlike copyright enforcement in court, where decisions are made public, we know very little about the actual implementation of the notice-and-takedown regime.

44 Bar-Ziv & Elkin-Koren, 2018, p. 377.

45 *UMG Recordings, Inc. v. Veoh Networks Inc.*, 665 F. Supp. 2d 1099, 1110 (C.D. Cal. 2009) at 1111: "UMG has not established that the DMCA imposes an obligation on a service provider to implement filtering technology (…)." However, some service providers have undertaken measures that exceed their legal obligations under the notice-and-takedown regime and voluntarily offer additional enforcement measures to copyright holders (e.g., YouTube's Content ID service). See also: Bridy, 2016, p. 192.

46 Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, *Official Journal* L 178, 17.7.2000.

47 See Arts. 12-14 of the Directive on electronic commerce.

EU rules were modeled on the DMCA; however, they differ from the US safe harbor in two ways. First and most importantly, the directive's hosting provision governs all claims related to user-generated content, not just copyright. These claims may be derived from private law, in the form of, e.g., copyright infringement or defamation, as well as from criminal law, in the form of, e.g., incitement to violence or hate speech. Second, the notice and takedown mechanism is prescribed by a directive that allows for certain flexibility within national legislators and has resulted in 27 harmonized, albeit not identical, national legal regimes in EU member states.[48] The e-commerce directive additionally prohibits the imposition of general obligations on hosts that are protected by a safe harbor to monitor the information which they transmit or store, or to actively seek out facts or circumstances indicating illegal activity.[49]

As already noted in US case law, the expeditious removal of content may be (mis) used for censorship purposes. For that reason, the Court of Justice of the European Union in the *Promusicae* case[50] clarified that in transposing the directives and implementing the transposing measures "the Member States must (…) take care to rely on an interpretation of the directives which allows a fair balance to be struck between the various fundamental rights protected by the Community legal order."[51] This 'fair balance' doctrine was also accepted and further developed by the ECtHR, particularly in the decisions *Delfi v. Estonia*[52] and *MTE v. Hungary*.[53] Both cases concerned online hosts' liability for allegedly defamatory content posted by anonymous users in the comment sections below news articles published by the platforms. In *Delfi v. Estonia,* the ECtHR listed four specific factors to guide the balancing process: (1) the context of the comments, (2) the measures applied by the platform in order to prevent or remove the comments, (3) the liability of the actual authors of the comments as an alternative to the platform's liability, and (4) the consequences of the domestic proceedings for the platform.[54] In *MTE v. Hungary*, the Court added a fifth factor: the consequences of the comments for the victim.[55] In applying these factors to the two cases, the ECtHR came to two opposite conclusions. In *Delfi v. Estonia*, the comments were qualified as hate speech and incitement to violence. Thus, the imposition of liability on the hosting provider struck a fair balance and therefore did not entail a violation of the right to freedom of expression. However, in *MTE v. Hungary*,

48 Before Brexit – 28.

49 Art. 15 of the Directive on electronic commerce.

50 CJEU, case C-275/06, *Productores de Música de España (Promusicae) v Telefónica de España SAU* [2008] 2
CMLR 465.

51 *Ibid,* para 68. Note: Rights derived from international law are referred to as human rights, while rights derived from domestic national constitutional law, as well as from European law, are referred to as fundamental rights.

52 ECtHR, *Delfi v. Estonia*, 16 June 2015.

53 ECtHR, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* (hereinafter: *MTE v. Hungary*), 2 February 2016.

54 ECtHR, *Delfi v. Estonia*, para. 142.

55 ECtHR, *MTE v. Hungary,* paras. 68–69.

the Court characterized the comments as merely offensive and concluded that the liability imposed on the intermediaries for their dissemination violated the right to freedom of expression. Although the fair balance doctrine remains somewhat unclear at present, it allows for much needed flexibility in the area of intermediary liability.

As our analysis has shown, EU legislation initially limited the action expected of the intermediary to only one possibility—takedown—which applied horizontally, i.e., to all areas of law in which intermediary liability arises as a potential issue. However, the Directive on Copyright in the Digital Single Market,[56] adopted in 2019, made a subtle variation from the notice and takedown mechanism to the more flexible notice and action mechanism. Article 17 of the directive regulates 'online content-sharing service providers' (OCSSPs). These are defined as platforms with a profit-making purpose that store and give the public access to a large amount of user-uploaded works/subject matter, which they organize and promote. This includes well-known platforms like YouTube and Facebook, as well as any type of user-upload platform that fits this broad definition and is not expressly excluded, as is the case with electronic communication services, providers of business-to-business cloud services and cloud services, online marketplaces, not-for profit online encyclopedias (e.g., Wikipedia), not-for-profit educational and scientific repositories, and open source software developing and sharing platforms. The directive states that OCSSPs carry out acts of communication to the public when they give access to works/subject matter uploaded by their users. As a result, these platforms become directly liable for their users' uploads. They are also expressly excluded from the hosting safe harbor for copyright relevant acts previously available to many of them under the e-commerce directive. Consequently, the platforms have two possibilities to avoid direct liability. First, they could obtain authorization to communicate or make the user-uploaded content available. However, it seems almost impossible to obtain authorization for all user-uploaded content. Consequently, OCSSPs will have to rely on the second possibility, which allows them to avoid liability if they meet a number of cumulative conditions. They must demonstrate that they have: (1) made best efforts to obtain an authorization, (2) made best efforts to ensure the unavailability of specific works for which the right holders have provided them with the relevant and necessary information, and (3) acted expeditiously, subsequent to notice from right holders, to take down infringing content and made best efforts to prevent its future upload. These conditions have been criticized in legal theory,[57] especially the second condition, which appears to impose an upload filtering obligation, and the third condition, which introduces both a notice and takedown mechanism (already

56 Directive (EU) 2019/790 of the European Parliament and of the Council on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, *Official Journal* L 130, 17.5.2019.

57 See for example: Quintais, 2020, pp. 28–41.

prescribed by the e-commerce directive) and a notice and stay down (or re-upload filtering) obligation.

In the interest of freedom of speech, the EU legislator created a special regime for certain copyright exceptions and limitations (quotation, criticism, caricature, review, parody, and pastiche).[58] However, existing content recognition technologies are not sophisticated enough, which could easily result in lawful uses of copyrighted works being blocked.

By adopting the Directive on Copyright in the Digital Single Market, the EU started a transition toward a 'vertical' approach to intermediary liability. This new approach can also be detected in new European legislation aimed at introducing a number of measures to prevent the misuse of Internet hosting services for the dissemination of texts, images, sound recordings, or videos that incite, solicit, or contribute to terrorist offenses. The regulation on addressing the dissemination of terrorist content online[59] is designed to establish binding, uniform rules that will, above all, ensure the swift removal of terrorist online content.[60] The regulation contains a uniform definition of terrorist online content, in line with EU fundamental rights protection. Service providers will have to remove terrorist content or disable access to it in all EU member states as soon as possible and in any event within one hour after they have received a removal order from a competent authority in an EU member state. Material disseminated for educational, journalistic, artistic, or research purposes, or that aims to prevent or counter terrorism will not be considered 'terrorist content;' this also includes content expressing polemic or controversial views in a public debate. The regulation includes effective remedies for both users whose content has been removed and service providers to submit a complaint.

The EU legal framework for social networks (in a broad sense) has also expanded with the latest review of the Audiovisual Media Services Directive (hereinafter 'AVMS Directive').[61] The AVMS Directive defines a 'video-sharing platform service' as a service where (i) the principal purpose of the service or of a dissociable section thereof or an (ii) essential functionality of the service is devoted to providing programmes, user-generated videos, or both, to the general public, for which the

58 Art. 17 and § 70 of the preamble of the Directive on the Digital Single Market.

59 Regulation (EU) 2021/784 of the European Parliament and of the Council on addressing the dissemination of terrorist content online, *Official Journal* L 172, 17.5.2021.

60 The removal of content is not the only activity that hosting service providers should undertake. According to the Proposal, providers should impose specific 'proactive measures' (see Art. 6 of the Proposal), although they do not have a general monitoring obligation. The Proposal states that in light of the particularly grave risks associated with the dissemination of terrorist content, the decisions adopted on the basis of the Regulation could, in fact, derogate from the prohibition of general monitoring set in the e-commerce directive. For an in-depth analysis of the Proposal, see: Kuczerawy, 2018, pp. 1–17.

61 Directive 2010/13/EU of the European Parliament and of the Council on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services, *Official Journal* L 095, 15.4.2010; L 263, 6.10.2010; L 303, 28.11.2018.

video-sharing platform provider does not have editorial responsibility. The service must be made available by means of an electronic communications network and the organization of the service determined by the video-sharing platform provider, including by automatic means or algorithms. The AVMS Directive states that in order for the provision of audiovisual content to constitute an 'essential functionality' of the service, such content must not be 'merely ancillary to, or a minor part of' the activities of the service. The European Commission's Guidelines on video-sharing platforms[62] set out several indicators that national authorities should consider, which can be grouped into four main categories: (1) the relationship between the audiovisual content and the main economic activities of the service; (2) quantitative and qualitative relevance of the audiovisual content available on the service; (3) monetization of, or revenue generation, from the audiovisual content; and (4) the availability of tools aimed at enhancing the visibility or attractiveness of the audiovisual content. Consequently, social media services can constitute video-sharing platform services and would fall within the scope of the AVMS Directive if they meet the relevant criteria.[63] The European Commission acknowledges that social media services have become an important medium by which users (particularly young people) access audiovisual content, and both the AVMS Directive and the Guidelines emphasize that because many social media services (i) compete for the same audiences and revenues as audiovisual media services and (ii) have a considerable impact, they must comply with the same regulations where they meet the relevant criteria.[64]

Although the AVMS Directive explicitly states that the e-commerce directive's 'safe harbor' provisions remain applicable, it requires member states to ensure that video-sharing platform providers operating within their respective jurisdictions take 'appropriate measures' to protect: (1) minors from programmes, user-generated videos and audiovisual commercial communications which may impair their physical, mental or moral development; (2) the general public from programmes, user-generated videos and audiovisual commercial communications containing incitement to violence or hatred directed against a group of persons or a member of a group; (3) the general public from programmes, user-generated videos and audiovisual commercial communications containing content the dissemination of which constitutes an activity which is a criminal offence under Union law, namely public provocation to commit a terrorist offence, offences concerning child pornography and offences concerning racism and xenophobia.[65] What constitutes an 'appropriate measure' is to be determined in light of the nature of the content in question, the harm it may cause, the characteristics of the category of persons to be protected as well as the

62 Communication from the Commission Guidelines on the practical application of the essential functionality criterion of the definition of a 'video-sharing platform service' under the Audiovisual Media Services Directive 2020/C 223/02 C/2020/4322, *Official Journal* C 223, 7.7.2020.

63 Services such as YouTube, as well as audiovisual content shared on social media services, such as Facebook, are covered by the revised AVMS Directive.

64 AVMS Directive, recital 4.

65 *Ibid*, art. 28b, para. 1.

rights and legitimate interests at stake, including those of the video-sharing platform providers and the users that created or uploaded the content, as well as the general public interest.[66]

The EU's interest in regulating online intermediaries was further demonstrated in late 2020, when the European Commission submitted a new legislative proposal to the European Parliament and European Council. The package consists of proposals of two regulations: the Digital Services Act[67] and the Digital Markets Act.[68] In the context of freedom of expression, the Digital Services Act is meant to improve the existing content moderation mechanisms. The Act will apply to online intermediaries ranging from cloud services and messaging services to marketplaces, Internet providers, and social networks. Further to this, specific due diligence obligations will apply to hosting services and online platforms, which are a subcategory of hosting services. The platforms will be required to disclose to regulators how their algorithms work, how decisions to remove content are taken, and the way advertisers target users. The Digital Services Act will create stronger public oversight of online platforms, particularly for platforms that reach more than 10% of the EU's population. Some of the measures proposed by the European Commission are: (1) measures to counter illegal goods, services or content online, such as a mechanism for users to flag such content and for platforms to cooperate with 'trusted flaggers;' (2) new obligations on traceability of business users in online market places, to help identify sellers of illegal goods; (3) effective safeguards for users, including the possibility to challenge platforms' content moderation decisions; (4) transparency measures for online platforms on a variety of issues, including on the algorithms used for recommendations; (5) obligations for very large platforms to prevent the misuse of their systems by taking risk-based action and through independent audits of their risk management systems; (6) access for researchers to the largest platforms' key data, in order to understand how online risks evolve; (7) oversight structure to address the complexity of the online space. We shall not further analyze the proposed rules, given that they could (and most probably will) be modified during the legislative process that has just started.

1.5. Social networks' internal rules on content moderation

In 1997, the US Government explicitly supported self-regulation as the primary mechanism for regulating the Internet in its report 'Framework for global electronic commerce,' stating that:

66 *Ibid,* art. 28b, para. 3.

67 Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final.

68 Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM/2020/842 final.

> (…) governments should encourage industry self-regulation wherever appropriate and support the efforts of private sector organizations to develop mechanisms to facilitate the successful operation of the Internet. Even where collective agreements or standards are necessary, private entities should, where possible, take the lead in organizing them.[69]

Today, more than twenty years later, we are witnessing different forms of online rules and regulations, such as terms of service,[70] privacy policies,[71] IP policies,[72] and community standards.[73] Although Internet platforms tend to present these rules as users' democratic participation in their services and may occasionally seek public feedback, they actually reflect the asymmetric relationship between platforms and users. More accurately, these rules are made and closely enforced by corporate entities and are far from the 'self-governance utopia' of the 1990s.

Further to the rules' lack of democratic legitimacy, the internal content moderation mechanisms demonstrate a striking transparency deficit. Due to the extreme volume of content posted online, these mechanisms are increasingly being applied automatically by way of artificial intelligence (AI), (almost) without any human interference. Automatic detection and filtering technologies are becoming essential tools in the fight against illegal online content. Indeed, many large platforms are now making use of some form of matching algorithms based on a range of technologies, from metadata filtering to hashing and fingerprinting content. However, the asymmetry of AI is even more problematic, since the user only sees the results of its individual decisions and has no access to accurate information about the input that determined a particular output.[74] Moreover, bias may be introduced into machine learning processes at various stages, including during algorithm design. Users have no information regarding the design or instructions the platforms input into the machine, and it could easily be a source of biases and over-removal.[75]

In its 2018 Recommendation on Measures to Effectively Tackle Illegal Content Online, the European Commission endorsed the provision of effective and appropriate safeguards to ensure that decisions taken concerning the removal of content are accurate and well-founded. In the Commission's view, such safeguards should consist, in particular, of human oversight and verification where appropriate and, in any event, where a detailed assessment of the relevant context is required in order to determine whether or not the content is to be considered illegal.[76] Moreover, if

69 White House, *The Framework for Global Electronic Commerce*, 1997. See: https://bit.ly/3lDsnnm.
70 See, for example, Twitter Terms of service, https://twitter.com/en/tos.
71 See, for example, Instagram Data policy, https://help.instagram.com/519522125107875.
72 See, for example, YouTube Copyright policy, https://bit.ly/3lDdT7a.
73 See, for example, Facebook Community standards, https://www.facebook.com/communitystandards/.
74 Castets-Renard, 2020, p. 23.
75 *Ibid.*
76 Recommendation on measures to effectively tackle illegal content online, C(2018) 1177 final, § 20.

the proposed Digital Services Act is adopted, intermediary service providers will be required to provide terms and conditions that include information about any restrictions that they impose on the use of their service in respect of information provided by the service recipients. That information will have to include information about any policies, procedures, measures, and tools used for the purpose of content moderation, including algorithmic decision making and human review.[77]

Finally, once a decision on content removal is reached, pursuant to the social network's internal rules, it is usually impossible to challenge. In most cases, there is no judicial review available when platforms take action against content or activity that violates their community standards or terms of service. Although some litigants are testing the limits of this obstacle before the US courts, since most Big Tech companies are headquartered in the United States, they have not yet prevailed.[78] However, in some other jurisdictions the courts have recognized that users have remedies against platforms that wrongfully delete content. In Germany, for instance, the courts have long applied the *Drittwirkung* doctrine, which recognizes that public law values influence private rights. On several occasions, the courts held that, under the *Drittwirkung* doctrine, Facebook must respect fundamental rights when it determines whether to delete content pursuant to its terms of service.[79]

There are numerous examples social media platforms' clear mistakes or at least questionable content removal decisions. For example, in 2016, Facebook, under its child pornography policy, blocked the sharing of the iconic 'Napalm Girl' photo depicting a young Vietnamese girl running naked and panicked from a napalm attack on her village. However, following widespread criticism from news organizations and media experts across the globe, Facebook reversed its decision.[80]

In response to longstanding criticism demanding user accountability, Mark Zuckerberg, CEO and founder of Facebook, the most popular social network,[81] announced in November 2018 that his company would create an independent governance and oversight committee by the close of 2019 to advise on content policy and listen to user appeals on content decisions.[82] In September 2019, Facebook published the Oversight Board Charter, a document that delineates the structural relationship between Facebook, the Oversight Board, and the Trust that ensures the Board's financial independence from Facebook.[83] The Oversight Board has between eleven and forty members; it will increase or decrease in size 'as appropriate.'[84] Members

77 Proposal of the Digital Services Act, art. 12, para. 1.

78 *Prager Univ. v. Google LLC*, No. 17-CV-06064-LHK, 2018 WL 1471939, at *14 (N.D. Cal. Mar. 26, 2018).

79 Bloch-Wehba, 2019, p. 77.

80 See for example: The Guardian, 'Facebook backs down from 'napalm girl' censorship and reinstates photo'. Available at: https://bit.ly/3EC00yO.

81 Per number of active users.

82 Mark Zuckerberg, 'A Blueprint for Content Governance and Enforcement', 15 November 2018. Available at: https://bit.ly/2XFrwLg.

83 Facebook Oversight Board Charter. Available at: https://bit.ly/3tZgagF.

84 Facebook Oversight Board Charter, art. 1. The names of the first twenty members were announced in May 2020.

of the Oversight Board must possess and exhibit a broad range of knowledge, competencies, diversity, and expertise, and must have demonstrated experience deliberating thoughtfully as an open-minded contributor on a team, be skilled at making and explaining decisions, and have familiarity with matters relating to digital content and governance, including free expression, civic discourse, safety, privacy, and technology.[85] The Charter also instructs the Board to split into subsections, termed panels, when reviewing cases. Each panel has to contain at least one member from the region where the case arose.[86]

Excluding content that was removed in compliance with local laws[87] and requiring following an exhaustion of appeals through Facebook, a request for review can be submitted to the Board by either the original poster of the content or a person who previously submitted the content to Facebook for review.[88] Consequently, the Oversight Board has the authority to review not only content that has been removed (original poster of the content) but content that is kept up (person who previously submitted content for review). However, the Facebook Oversight Board bylaws create many exceptions to the Board's scope of review. As established at the Board's launch,[89] only single-object removals of organic content posted on Facebook and Instagram are eligible for review.[90] Within that, content decisions 'pursuant to legal obligations,' including those concerning intellectual property, the Facebook marketplace, fundraisers, Facebook dating, messages, and spam, are out of the scope.[91]

1.6. Social networks between proclaimed neutrality and value-based decisions

Following a brief period of euphoria about the possibility that social networks might facilitate global democratization, there is now widespread concern in many segments of society that social networks may instead be undermining democracy. Their specific role in a digital society does not easily fit into any of the existing categories. They cannot be qualified as 'speakers,' as they do not publish their own content, nor do they associate themselves with the content their users publish. They cannot be qualified as a traditional 'editor' either, as they do not initiate or

85 *Ibid.*
86 *Ibid.*
87 *Ibid,* art. 7.
88 *Ibid,* art. 2.
89 The type of content eligible for review can be broadened in time. For a critical assessment, see: Klonick, 2020, p. 2465 *et seq.*
90 'Organic content' is content posted by users, contrary to commercial advertising. 'Single-object' refers to a post containing a photo, video, or status message. 'Complex object' is a user profile, group, or page.
91 Facebook Oversight Board Bylaws, art. 2, § 1.2. See: https://www.oversightboard.com/sr/governance/bylaws.

commission the production of content. However, they do exercise certain editorial functions in the sense that they moderate the content their users post.[92]

The system that social networks have put in place to match users' expectations and self-regulate is indeed responsive, as demonstrated in our analysis. However, this system presents two major downsides that become more apparent over time. First, there is an evident loss of equal access to and participation in speech on these platforms.[93] Social networks are increasingly making their own choices regarding content moderation that give preferential treatment to some users over others, e.g., by designing algorithms in accordance with the network owner's preferences. Moreover, algorithms are often set to create perfect filtering in order to only show users content that meets their personal tastes. This may create a basically antidemocratic space in which people are shown things with which they already associate. As a number of social science researchers have rightfully noted,[94] although the rise of social media has made citizens much less dependent on television and traditional newspapers, this certainly does not mean that citizens have more control over the media environments in which they now operate. Media power has not been transferred to the public; instead, power has partly shifted to algorithmic selections operated by large digital platforms.

The second problem is that of accountability. Social networks should be open about their takedown rules and follow a consistent and transparent process. Under the current legal regime, the user is virtually powerless. Users are not sufficiently informed about the criteria social networks apply when moderating content. In most cases, the user cannot successfully challenge the platform's content moderation decisions either. Greater transparency in content moderation implies publication of the number of posts and accounts being removed, provision of a clear notice to users disclosing the reason for content removal, and human review of removal decisions undertaken by software.

2. Fake news as a global factor in the influence of social networks on the guarantees of freedom of speech and the truthfulness of information

In recent years, concerns about the societal consequences of the online spread of disinformation and propaganda have become widespread. New digital tools that allow anyone to easily spread political information to large numbers of Internet users can lead to a more pluralistic public debate, but they can also give a platform

92 Koltay, 2019, p. 189.
93 Klonick, 2017, p. 1665.
94 See for example: Poell and van Dijck, 2015, pp. 527–537.

to extremist voices and actors seeking to manipulate the political agenda in their own political or financial interest.[95] The problem of 'fake news' attracted substantial attention during the 2016 US presidential elections, after a series of events known as 'Pizzagate.' Namely, fake news publishers in North Macedonia circulated a false political conspiracy theory that former First Lady, Secretary of State, and presidential candidate Hillary Clinton and other prominent Democratic political figures were coordinating a child trafficking ring out of a Washington-based pizzeria by the name of Comet Ping Pong. This fake news was widely shared via social networks. In December 2016, a man who read the publication drove from North Carolina to Washington, DC and shot open a locked door at Comet Ping Pong pizzeria with his assault rifle.[96]

False statements of fact typically published on websites and disseminated via social networks for profit or social influence are usually referred to as fake news, rumors, counter-knowledge, disinformation, post-truths, alternative facts, or simply lies. Although this phenomenon is omnipresent, it is rarely defined in legal documents (Section 2.1). More recently, the concept of 'deep fakes' has been introduced (Section 2.2). The creation and/or dissemination of fake news may result in civil, criminal, or administrative liability for Internet users. Moreover, social networks have adopted their own internal rules aimed at combatting the dissemination of fake news (Section 2.3). Some governments and non-governmental organizations, either on their own or in collaboration with social networks, have introduced media literacy initiatives as an alternative approach to combatting fake news (Section 2.4).

2.1. The concept of fake news

The UK Collins Dictionary named 'fake news' the 2017 'word of the year.' According to the dictionary, usage of the phrase indicating "false, often sensational, information disseminated under the guise of news reporting" increased by 365% since 2016.

The two defining characteristics used to identify different types of fake news are, first, whether the author intends to deceive readers and, second, whether the motivation for creating or disseminating the fake news is financial.[97] By applying these two criteria, one could differentiate among at least four types of fake news. The first type is satire, that is, a news story that does not intend to deceive, although it purposefully contains false content, and is generally motivated by non-pecuniary interests, though financial benefit may be a secondary goal. The second type of fake news is a hoax, which is a news story with purposefully false content where the author intends to deceive readers into believing incorrect information and that is

95 Tucker et al., 2018, p. 15.

96 BBC, 'The saga of Pizzagate: The fake story that shows how conspiracy theories spread'. Available at: https://bbc.in/39tv59i.

97 Verstraete et al., 2017, p. 6.

financially motivated. Typically, creators of hoaxes do not have political or cultural motivations that drive the production of their fake news stories. The third type is propaganda, which is news or information with purposefully biased or false content where the author intends to deceive readers and that is motivated by promoting a political cause or point of view, regardless of financial reward. Fourth, 'trolling' presents news or information with biased or fake content where its author intends to deceive readers and is motivated by an attempt to derive personal humorous value (the lulz).[98] The term 'fake news' has a distinctively negative connotation, which is why the general public's understanding is usually limited to the second and third types of activities (i.e., hoax, propaganda).

Given its complexity and the different perceptions, the term 'fake news' is less employed in legal doctrine and legal documents in recent years. Instead, it is being replaced by the term 'disinformation.' This is particularly the case in the EU in the context of recent European Commission initiatives. Specifically, in 2018, the European Commission set up a high-level expert group on fake news and online disinformation to advise the Commission on establishing the scope of the disinformation phenomenon, defining the roles and responsibilities of relevant stakeholders, and formulating recommendations. The expert group released its final report[99] only a few months later. This was followed by the European Commission's Communication titled 'Tackling Online Disinformation: A European Approach.'[100] In September 2018, the European Commission published the Code of Practice on Disinformation (hereafter, 'the Code').[101] The Code represents a voluntary, self-regulatory mechanism agreed upon by representatives of online platforms, social networks, advertisers, and the advertising industry. The Code employs the term 'disinformation,' defined as 'verifiably false or misleading information' that is both "created, presented and disseminated for economic gain or to intentionally deceive the public" and may cause public harm, intended as "threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security."[102] The term does not cover misleading advertising, reporting errors, satire and parody, or clearly identified partisan news and commentary.[103] Moreover, disinformation as defined here includes forms of speech that fall outside already illegal forms of speech, notably defamation, hate speech, incitement to violence, etc., but can nonetheless be harmful.[104]

98 'Lulz' is a typographical subversion of the word 'lol,' meaning to 'laugh out loud.'

99 European Commission, Final report of the High level expert group on fake news and online disinformation, 'A multi-dimensional approach to disinformation', 2018. See: https://bit.ly/3zt3bF2.

100 European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions, 'Tackling online Disinformation: a European Approach', COM(2018) 236 final.

101 European Commission, Code of practice on disinformation, 2018. See: https://bit.ly/39rdpey.

102 *Ibid*, preamble, p. 1.

103 *Ibid.*

104 Final report of the High level expert group on fake news and online disinformation, p. 10.

The reason the EU seems to prefer the term 'disinformation' to 'fake news' is explained in the Final Report of the High Level Expert Group on Fake News and Online Disinformation. First, the latter term is considered to be inadequate to capture the complex problem of disinformation, which involves content that is not actually or completely 'fake' but is rather fabricated information blended with facts and practices that go well beyond anything resembling 'news' to include some forms of automated accounts used for astroturfing, networks of fake followers, fabricated or manipulated videos, targeted advertising, organized trolling, visual memes, and much more. Second, the term 'fake news' has been appropriated by some politicians and their supporters, who use it to dismiss coverage that they find disagreeable.[105]

2.2. The concept of deep fakes

'Deep fakes' are face-swapping technologies that enable the quick creation of fake images or videos that appear very realistic. Deep fake technology can also be used to create 'voice clones,' usually of public figures. Typically, deep fakes rely on artificial neural networks, which are computer systems that recognize patterns in data. Developing a deep fake photo or video involves feeding hundreds or thousands of images into the artificial neural network in order to 'train' it to identify and reconstruct patterns. Coinage of the term 'deep fakes' is attributed to a Reddit[106] user called 'deepfakes,' who published several videos in which famous actresses' faces were swapped into pornographic videos in late 2017.[107] The increased availability of deep fakes, especially through apps, raises a number of legal, social, and ethical questions. Indeed, their very existence is blurring the line between what is true and what is fake.

Legal theory distinguishes among four main types of deep fakes.[108] The first type is deep fake pornography, for which technology is used either to create celebrity deep fakes or revenge porn. Celebrity deep fakes refer to content where celebrity images are superimposed on the bodies of individuals engaged in sexual acts. Revenge porn is created by persons seeking revenge for terminated relations. The second type of deep fake comprises fake photos or videos created during a political campaign. This type of deep fake can have significant negative consequences for democratic processes, as deep fakes can target certain individuals' reputation or portray fake events. The third type of deep fake comprises fake photos, videos, or voices created for commercial purposes. For example, the technology can be used to translate a video by enabling the recorded person to 'speak' in different languages.

105 *Ibid.*

106 Reddit is a website comprising user-generated content (including photos, videos, links, and text-based posts) and discussions of this content in what is essentially a bulletin board system.

107 The New York Times, 'Here come the fake videos, too'. See: https://nyti.ms/3AtUH1X.

108 See for example: Meskys et al., 2020, pp. 24–31.

Finally, the fourth type could be referred to as a creative deep fake. This category comprises fake content created purely for creative purposes, usually as parody or satire.

2.3. Legal framework for combatting the creation and dissemination of fake news

The creation and/or dissemination of fake news may result in civil, criminal, or administrative liability for Internet users. Further to these 'traditional' legal instruments, legislators in certain jurisdictions have adopted specific legislative acts aimed at combatting the creation and dissemination of fake news. We will analyze in further detail the existing legal framework related to fake news in the United States, on the one hand, and in the EU and its member states, on the other hand. Moreover, social networks have adopted their own internal rules aimed at combatting the dissemination of fake news.

2.3.1. US law

Fake news creators and/or disseminators are frequently sued by private individuals or businesses seeking to collect monetary damages or injunctive relief in civil law proceedings. The most frequent claim invoked against fake news creators and/or disseminators is the common law tort of defamation.[109] In the United States, false publications of fact concerning a public figure (e.g., a government official) are actionable only if the publisher acted with actual malice, i.e., either with knowledge of the statement's falsity or reckless disregard for the same. However, strictly private figures do not need to prove actual malice; they are only required to prove that the defamatory statements were published with negligence. If we define fake news restrictively, so as to include only intentional or knowingly false statements, it is reasonable to conclude that such statements would satisfy the requirements for defamation claims. However, fake news in a broad sense need not always satisfy these requirements. For example, a satire or parody is actionable only if it could be reasonably understood to describe actual facts or events, which is typically not the case. Finally, it should be recalled that Section 230 of the Communications Decency Act of 1996 protects online publishers[110] from defamation claims in situations where another Internet user provided the information.

After defamation, intentional infliction of emotional distress (IIED) is a common law tort that is regularly alleged against fake news creators and/or disseminators under state law. IIED occurs when a person intentionally or recklessly engages in

109 Defamation is the communication of a false statement of fact that harms another person's reputation or character. Spoken (unrecorded) defamation is referred to as slander, while defamatory statements that are written or otherwise recorded are known as libel.

110 However, it does not protect the original author of a defamatory or otherwise tortious publication.

extreme or outrageous behavior that causes another person to suffer severe emotional distress. Unlike defamatory statements, which may be actionable for simply being harmful and false, statements supporting IIED claims must be "so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious, and utterly intolerable in a civilized community."[111] Consequently, particularly extreme fake news content remains susceptible to IIED claims, especially when involving non-public figures.

Moreover, creating fake news content could easily violate a third party's intellectual property rights, typically a copyright or trademark right. The creators of text, photographs, videos, and other original works of authorship are granted exclusive rights to reproduce, distribute, display, and create derivative works from such content. Consequently, creators and/or disseminators of fake news content using third-party materials have to seek the copyright owners' permission (unless the work is in the public domain or the doctrine of fair use applies). In addition, the creators of fake news content should refrain from using third-party trademarks or logos that may confuse consumers as to the origin of products, since the Lanham Act and state unfair competition law prohibit trademark infringements and false representations of fact in commercial advertising that misrepresent the nature or characteristics of another's goods, services, or commercial activities.[112] Creators and/or disseminators of fake news content may also be sued for the violation of the right of publicity, i.e., respect for a person's name and likeness, which most US states recognize.[113] The right of publicity grants an individual the right to control the *commercial* use of their identity.

In addition to civil law liability, fake news creators and/or disseminators may be accused of crimes or the violation of other specific regulations. For example, the Federal Trade Commission (FTC) is given broad discretion to investigate questionable trade practices and take appropriate enforcement action. Entities found to have engaged in consumer fraud or deception can be permanently enjoined by a court from continuing such conduct in the future. They may also be ordered to pay civil penalties and provide consumer redress.[114] Further to this, criminal libel statutes exist in several US states and territories.[115] The elements of criminal libel are similar to the elements of civil defamation. Criminal libel consists of defamation of an individual (or group) made public by a printing or writing. The defamation must

111 Restatement (Second) of Torts § 46 cmt. d (Am. Law Inst. 1965). For a critical analysis of IIED see: Fraker, 2008, pp. 983–1026.

112 15 U.S.C. § 1125(a).

113 See for example: N.Y. Civ. Rights Law § 50.

114 Within the FTC is the Bureau of Consumer Protection, which is designed to protect consumers from deceptive or unfair business practices. The Bureau of Consumer Protection focuses on protecting consumers' privacy, fighting identity theft, regulating advertising and marketing practices, regulating business practices in the financial industry, and protecting US citizens from telemarketing fraud.

115 For example, in Florida (see: Chapter 836 of the Florida Statutes).

tend to excite a breach of the peace or damage the individual (or group) in reference to their character, reputation, or credit.[116]

Finally, in October 2017, Congress announced a bill that would require digital platforms with at least 50,000,000 monthly visitors to maintain a public file of all electioneering communications purchased by a person or group who spends more than $500.00 in total on ads published on their platform. This file must contain a digital copy of the advertisement, a description of the audience the advertisement targets, the number of views generated, the dates and times of publication, the rates charged, and the purchaser's contact information. The bill, called the Honest Ads Act, was introduced by US senators Mark Warner, Amy Klobuchar, and Lindsey Graham, with the aim of preventing foreign interference in future elections and improving the transparency of online political advertisements.[117] The proposed legislation addresses a loophole in the existing campaign finance laws that regulate television and radio ads, but not Internet ads. The Honest Ads Act would help close that gap by subjecting Internet ads to the same rules as television and radio ads.

2.3.2. European Union and its member states

The problem of disinformation on the Internet is a source of growing concern for EU policymakers. As previously mentioned, in September 2018, the European Commission published the Code of Practice on Disinformation, which is a voluntary, self-regulatory mechanism agreed upon by representatives of online platforms, social networks, advertisers, and the advertising industry. The Code observes that social networks facilitate the dissemination of disinformation, impacting a broad segment of actors in the ecosystem. For this reason, all stakeholders have roles to play in countering the spread of disinformation.[118] The Code considers advertising and monetization incentives as leading to behaviors such as misrepresentations about oneself or the purpose of one's properties.[119] In response, the Code's signatories have committed to deploying policies and processes to disrupt such incentives. The signatories have acknowledged, in particular, that there is a need to significantly improve the scrutiny of ad placements.[120] All parties involved in the online advertising market need to work together to improve transparency across the ecosystem. This means that they should effectively scrutinize, control, and limit the placement of advertising on accounts and websites belonging to purveyors of disinformation.[121] The signatories, moreover, should make commercially reasonable efforts to ensure that they do not accept remuneration from or promote accounts and websites that

116 Brenner, 2007, p. 714.
117 The full text of this legislative proposal is available here: https://bit.ly/2XBWKCA.
118 Code of Practice on Disinformation, p. 1.
119 *Ibid,* p. 5.
120 *Ibid,* p. 4.
121 *Ibid,* p. 4.

consistently misrepresent information about themselves.[122] The Code acknowledges the need to ensure transparency in the area of political and issue-based advertising. In particular, such transparency means that users should be able to understand why they have been targeted for a given advertisement.[123]

Some of the self-regulatory standards introduced by the Code are reflected in the European Commission's proposal of the Digital Services Act, published in December 2020.[124] The Act is supposed to impose greater transparency obligations for platforms in the field of targeted advertising, amongst other requirements in the field of content regulation. Penalties for violations of the rules include fines of up to 6% of a company's annual income.[125] In the field of online advertising, the European Commission has proposed rules that would give online platform users immediate information about the sources of the ads they see online, including granular information about why an individual has been targeted with a specific advertisement.[126] Moreover, very large online platforms[127] that display advertising on their online interfaces will have to compile and make publicly available through application programming interfaces a repository containing the following information: (1) the content of the advertisement; (2) the natural or legal person on whose behalf the advertisement is displayed; (3) the period during which the advertisement was displayed; (4) whether the advertisement was intended to be displayed specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose; (5) the total number of recipients of the service reached and, where applicable, aggregate numbers for the group or groups of recipients whom the advertisement targeted specifically. The information will have to remain publicly available until one year after the last time the advertisement was displayed on their online interfaces.[128]

Several EU member states have complemented the EU's current self-regulatory approach, which is best demonstrated in the Code of Practice on Disinformation, with its mandatory rules and harsher sanctions for non-compliance. Germany reacted first, although its reaction was directed more toward hate speech than fake news. In September 2015, the German Minister of Justice first initiated a task force composed of representatives of the service providers Facebook, Twitter, and Google (with respect to its service YouTube), and several nongovernmental organizations (NGOs) to jointly fight illegal speech. The self-regulatory measures they agreed upon included user-friendly notification mechanisms, an immediate review of notified content for

122 *Ibid.*

123 *Ibid,* p. 5.

124 Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final.

125 *Ibid,* arts. 42 and 59.

126 *Ibid,* art. 24.

127 Online platforms that provide their services to a number of average monthly active service recipients in the Union equal to or higher than 45 million.

128 Digital Services Act*,* art. 30.

compatibility with German law (within 24 hours of notification), adequate responses to illegal hate speech including the blocking of access to domestic users without undue delay, and transparent notice and takedown policies.[129] In spite of leading social networks' willingness to implement this self-regulatory mechanism, Germany proceeded with the adoption of harsher mandatory rules against illegal content online. In 2017, German Parliament adopted the Law Improving Law Enforcement on Social Networks (NetzDG).[130] This federal law aims at improving law enforcement regarding social networks by calling 'telemedia service providers'[131] to account regarding acting on online speech that is punishable under domestic criminal law. The NetzDG applies to all telemedia service providers that, for profit-making purposes, operate Internet platforms designed to enable users to share any content with other users or make such content available to the public.[132] Social network operators with at least two million registered users within Germany are required to implement an effective, transparent complaints management infrastructure and have the duty to compile reports on complaints management activity.[133] The law distinguishes between content that is manifestly illegal and that which is illegal. Manifestly illegal content must be deleted or removed within 24 hours of receiving a complaint, while for merely illegal content, a period of seven days is granted for action.

As neither hate speech nor the dissemination of fake news as such are statutory offenses under German criminal law, the NetzDG lists a catalogue of offenses considered to be illegal content requiring access blocking: (1) dissemination of propaganda material of unconstitutional organizations; (2) usage of symbols of unconstitutional organizations; (3) preparation of a serious violent offense endangering the State; (4) encouraging the commission of a serious violent offence endangering the state; (5) treasonous forgery; (6) public incitement to crime; (7) breach of the public peace by threatening to commit offense; (8) forming criminal or terrorist organizations; (9) incitement to hatred; (10) dissemination of depictions of violence; (11) rewarding and approving of offenses; (12) defamation of religions, religious and ideological associations; (13) distribution of child pornographic performances by broadcasting, media services or telecommunications services; (14) insult; (15) defamation; (16) violation of intimate privacy by taking photographs; (17) threatening the commission of a felony; and (18) forgery of data intended to provide evidence.[134]

129 Schmitz-Berndt and Berndt, 2018, p. 15.

130 *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*, Bundesgesetzblatt Teil 1 (BGB 1), n° 61, 7 September 2017.

131 Telemedia service providers are defined as electronic information and communications services, insofar as they do not provide telecommunications services, which consist of the transmission of signals via telecommunications networks, telecommunications-based services, or broadcasting services.

132 NetzDG, § 1(1).

133 *Ibid,* § 2-3.

134 *Ibid,* § 1(3).

Paradoxically, although the battle against fake news has been one of the main arguments to pass the NetzDG, the notion does not appear in the law itself.[135]

In November 2018, neighboring France adopted the Law Against the Manipulation of Information,[136] which targets the widespread and extremely rapid dissemination of fake news by means of digital tools, especially through the dissemination channels offered by social networks and media outlets influenced by foreign states. The law requires online platforms with more than five million unique users per month in France to adhere to the following conduct during the three months preceding general elections: (1) provide users with honest, clear, and transparent information about the identity and corporate address of anyone who paid to promote informational content related to a 'debate of national interest;' (2) provide users with honest, clear, and transparent information about the use of personal data in the context of promoting content related to a 'debate of national interest;' (3) make public the amount of payments received for the promotion of informational content when these amounts are above a certain threshold.[137] Moreover, the law provides that, during the three months preceding an election, a judge may order 'any proportional and necessary measure' to stop the deliberate, artificial, or automatic and massive dissemination of fake or misleading information online.[138] A public prosecutor, candidate, political group or party, or any person with standing can bring a fake news case before a judge, who must rule on the motion within 48 hours.[139] An interim judge will qualify the fake news, as defined in the 1881 Law on the Freedom of the Press, in accordance with three criteria: (1) the fake news must be manifest, (2) be disseminated deliberately on a massive scale, and (3) lead to a disturbance of the peace or compromise the outcome of an election.[140] Further to this, the Law Against the Manipulation of Information requires that online platform operators implement measures to prevent the dissemination of false information that could disturb public order or affect the validity of an election.[141] They must also establish an easily accessible mechanism for users to flag fake information, and they are required to submit a yearly report to the French Superior Council on Audiovisual (CSA)[142] detailing the measures they have taken to curb the dissemination of fake news.[143]

Italy also reacted to the online spread of disinformation by introducing a specific enforcement mechanism to combat fake news during the election period. In January

135 Schmitz-Berndt and Berndt, 2018, p. 21.

136 *Loi n° 2018–1202 relative à la lutte contre la manipulation de l'information,* Official Journal n°0297 of 23 December 2018. This 'ordinary law' is paired with the 'organic law' against the manipulation of information: *Loi organique n° 2018–1201 relative à la lutte contre la manipulation de l'information,* Official Journal n°0297 of 23 December 2018.

137 (Ordinary) law against the manipulation of information, art. 1.

138 *Ibid.*

139 *Ibid.*

140 Law on the freedom of press (*Loi du 29 juillet 1881 sur la liberté de la presse*), art. 27.

141 (Ordinary) law against the manipulation of information*,* art. 11.

142 *Conseil supérieur de l'audiovisuel.*

143 (Ordinary) law against the manipulation of information, art. 11.

2018, the minister of the interior introduced the Operating Protocol for the Fight Against the Diffusion of Fake News through the Web on the Occasion of the Election Campaign for the 2018 Political Elections.[144] General elections were scheduled for March 2018.[145] The protocol introduced a 'red button' reporting service where users "may indicate the existence of a network of content attributable to fake news." The Polizia Postale, a unit of the Italian State Police that investigates cybercrime, were tasked with reviewing reports and acting accordingly. The web portal allowed users to submit links to content and social networks (if they found the content on a social network), as well as further information. The portal also required users to provide their email address. The police then reviewed submissions with the aim of 'directing the next activity' for content that is 'manifestly unfounded and biased' or 'openly defamatory.' The police were supposed to carry out in-depth analysis using specific techniques and software in order to identify significant indicators allowing for the qualification, with maximum certainty, of the news as fake news (presence of official denials, false content already proven by objective sources, provenance of the alleged fake news from sources not accredited or certified, etc.). The Polizia Postale were also empowered to independently collect information "in order to identify early on the network of news markedly characterized by groundlessness and tendency that is openly defamatory." After reviewing the information, the authorities would pursue legal action if they determined that the content was unlawful. In cases where content was deemed to be false or misleading, but not unlawful, authorities would publish public denials.

The operating protocol contained references to defamation, which the Italian Penal Code defines as "injuring the reputation of an absent person via communication with others" and to which it attaches penalties of up to one year of imprisonment for members of the general public.[146] If the defamatory act or insult consisted of the allegation of a specific fact, the potential penalty increased to imprisonment for up to two years or a fine of 2,065 euros.[147] If committed by the press or otherwise publicly, violators could face penalties of at least 516 euros or imprisonment from six months to three years.[148] The penal code also provided for increased penalties for defamation against public officials. For example, the code imposed enhanced penalties of one to five years of imprisonment for criminal defamation of the president.[149] The Italian enforcement mechanism introduced in 2018 was criticized by the United Nations Human Rights Council (UN HRC) for failing to precisely define the type of

144 Press release: *Protocollo Operativo per il contrasto alla diffusione delle Fake News attraverso il web in occasione della Campagna elettorale per le Elezioni politiche 2018,* 18 January 2018. Available at: commissariatodips.it.

145 More on Italy's failed attempts to regulate 'fake news' prior to the adoption of the Operating protocol: Pollicino and Somaini, 2020, pp. 171–193.

146 Penal Code (*Codice Penale*), Official Journal n. 251/1930, art. 595.

147 *Ibid.*

148 *Ibid.*

149 *Ibid,* art. 278.

disinformation it targeted. The operating protocol aimed at combatting "manifestly unfounded and biased news, or openly defamatory content" left significant discretionary power to the police, according to the UN HRC special rapporteur on the promotion and protection of the right to freedom of opinion and expression.[150] Following widespread criticism, the authorities stopped enforcing the protocol.

2.3.3. Social networks' internal rules against fake news

Most social networks do not have a blanket rule against posting false material, but they do ban certain kinds of disinformation. Some allow specific types of false claims. For example, Facebook admits in its Community Standards that it does not totally ban fake news:

> Reducing the spread of false news on Facebook is a responsibility that we take seriously. We also recognize that this is a challenging and sensitive issue. We want to help people stay informed without stifling productive public discourse. There is also a fine line between false news and satire or opinion. For these reasons, we don't remove false news from Facebook but instead, significantly reduce its distribution by showing it lower in the News Feed.[151]

Twitter also stated that it is not addressing all false material:

> We are not attempting to address all misinformation. Instead, we prioritize based on the highest potential for harm, focusing on manipulated media, civic integrity, and COVID-19. Likelihood, severity and type of potential harm — along with reach and scale — factor into this.[152]

Social networks' internal rules against fake news are often vague and allow for a significant discretionary power as to whether the content will be blocked/permanently removed or not. TikTok, one of the newest social media, provides a good example of such rule ambiguity: "We do not permit misinformation that causes harm to individuals, our community, or the larger public regardless of intent."[153]

Social networks generally prohibit deep fakes, a specific type of manipulated content. For example, TikTok, Instagram, and Facebook explicitly prohibit AI-modified content: "Videos cannot be modified with AI tools in ways that are not apparent

150 Comments of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression on the Operating protocol for the fight against the diffusion of fake news through the web on the occasion of the election campaign for the 2018 political elections, p. 4. Available at: https://bit.ly/3CxdMRw.

151 Facebook Community Standards, § 20. See: https://www.facebook.com/communitystandards/false_news.

152 Tweeted on 'Twitter Safety' profile on 3 June 2020.

153 Tik-Tok Community Standards, section: Misinformation. See: https://bit.ly/3Cxdmul.

to an average person, and would likely mislead an average person to believe that a subject of the video said words that they did not say."[154] On the other hand, YouTube has more lenient rules regarding deep fakes: "Videos must not be technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm."[155]

Advertisers face stricter rules than 'ordinary users' on almost every platform. For example, Facebook only fact checks 'regular posts' (written by 'ordinary users') under special circumstances, while paid advertisements are always checked before being published. Paid ads need to comply with Facebook's advertising policies, which cover misinformation *inter alia,* but also with its Community Standards, which apply to regular posts as well. Under Facebook's advertising policies, ads that include claims debunked by third-party fact checkers or, in certain circumstances, by organizations with particular expertise, are prohibited. Advertisers that repeatedly post information deemed to be false may have restrictions placed on their ability to advertise on Facebook.[156] However, ads are rarely checked by human moderators. Instead, Facebook uses an algorithmic ad screening system. Similarly to Facebook, Twitter claims that it does not allow ads that are false, deceptive, misleading, defamatory, or libelous.[157]

2.4. Alternative approaches to combatting fake news

The self-regulatory approach, which social networks prefer, as well as the co-regulatory approach, which the EU favors, typically face several challenges. First, conflicts of interest may occur between the social networks' need to keep users engaged and monetize their engagement, and the public authorities' need to the safeguard the integrity of democratic processes. Second, the amount of content that has to be monitored is enormous, which necessarily implies the use of algorithmic content screening and consequently introduces possible errors in that process. Third, the efficiency of fact checking mechanisms is limited, as algorithms cannot be relied upon to control the extremely vast amount of online content. On the other hand, direct state-imposed regulation, which is preferred by certain European and non-European countries, focuses on illegal content, while ignoring many other variants of disinformation. Moreover, there is no commonly accepted definition of 'fake news,' which leaves significant discretionary power to enforcers.

Given that it has recently become increasingly difficult to recognize fake news and particularly deep fake materials, some alternative approaches to combatting disinformation have also been designed and implemented. Many governments and

154 Facebook Community Standards, § 21. Similar rules are adopted by other two networks.

155 YouTube Policies, section: Spam, deceptive practices, and scams policies. See: https://bit.ly/3tXuKW0.

156 Facebook Advertising Policies, § 13. See: https://www.facebook.com/policies/ads/.

157 Twitter Ads Policies. See: https://business.twitter.com/en/help/ads-policies.html.

NGOs have launched different media literacy initiatives, sometimes in collaboration with social network operators. Media literacy is usually defined as an informed, critical understanding of the prevalent mass media, and it involves examining the techniques and institutions involved in media production, as well as the ability to critically analyze media messages. One of the aspects of digital media literacy is the ability to recognize disinformation or partially false digital content.

The European Commission has also recognized that media literacy is a crucial skill for all European citizens, as it helps them to counter the effects of disinformation campaigns and the spreading of fake news through digital media. The revised AVMS Directive strengthens the role of media literacy. It requires EU member states to promote measures that develop media literacy skills.[158] The AVMS Directive also obliges video-sharing platforms to provide effective media literacy measures and tools. This is a crucial requirement due to the central role such platforms play in providing access to audiovisual content. Platforms are also required to raise users' awareness of these measures and tools.[159] Additionally, the European Commission has established a media literacy expert group that brings media literacy stakeholders together. This group meets annually to (1) identify, document and extend good practices in the field of media literacy; (2) facilitate networking between different stakeholders; and (3) explore ways of coordinating EU policies, support programmes and media literacy initiatives.[160]

An alternative approach to combatting fake news consists of fact checking projects oriented toward monitoring the factual accuracy of news, political statements, and interviews. Fact checking web portals offer counter-narratives to untrue and manipulated information. Facebook and Instagram have also established a fact checking program, in partnership with independent third-party fact checkers who are certified through the non-partisan International Fact-Checking Network (IFCN). The fact checking program, launched in 2016, enables fact checking partners to review content across both Facebook and Instagram, including organic and boosted posts. They can also review videos, images, links, and text-only posts.

158 AVMS Directive, art. 33a.

159 *Ibid,* art. 28b.

160 European Commission, Directorate-General for Communications Networks, Content and Technology, Mandate of the Expert Group on Media Literacy, 6 July 2016. Available at: https://bit.ly/39tv19y.

Bibliography

Balasubramani, V. (2016/2017) 'Online Intermediary Immunity Under Section 230', *The Business Lawyer,* 72(1), pp. 275–286.

Bar-Ziv, S. and Elkin-Koren, N. (2018) 'Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown', *Connecticut Law Review,* 50(2), pp. 339–385.

Bloch-Vehba, H. (2019) 'Global platform governance: Private power in the shadow of the state', *SMU Law Review,* 72(1), pp. 27–80.

Brenner, S.W. (2007) 'Should online defamation be criminalized?', *Mississippi Law Journal,* 76(3), pp. 705–787.

Bridy, A. (2016) 'Copyright's Digital Deputies: DMCA-Plus Enforcement by Internet Intermediaries' in Rothchild, J. A. (ed.) *Research Handbook on Electronic Commerce Law.* 1st Cheltenham: Edward Elgar, pp. 185–209.

Castets-Renard, C. (2020) 'Algorithmic content moderation on social media in EU law: Illusion of perfect enforcement', *University of Illinois Journal of Law, Technology & Policy*, 2020(2), pp. 283–323.

Durach, F. et al. (2020) 'Tackling Disinformation: EU Regulation of the Digital Space', *Romanian Journal of European Affairs,* 20(1), pp. 5–20.

ECtHR (2020) *Guide to Article 10 of the Convention – Freedom of expression.* Strasbourg: Council of Europe (online edition).

Fraker, R. (2008) 'Reformulating Outrage: A Critical Analysis of the Problematic Tort of IIED', *Vanderbilt Law Review*, 61(3), pp. 983–1026.

Frosio, G. Geiger, C. (2021) 'Taking fundamental rights seriously in the Digital Services Act's platform liability regime'. Available at: http://cyberlaw.stanford.edu/publications/taking-fundamental-rights-seriously-digital-service-act%E2%80%99s-platform-liability-regime (Accessed: 15 April 2021).

Klonick, K. (2017) 'The new governors: the people, rules and processes governing online speech', *Harvard Law Review,* 131(6), pp. 1598–1670.

Klonick, K. (2020) 'The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression', *Yale Law Journal,* 129, pp. 2418–2499.

Koltay, A. (2019) *New media and freedom of expression: rethinking the constitutional foundations of the public sphere*. London: Hart Publishing.

Kuczerawy, A. (2018) 'The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risksfor freedom of expression', *CDT Working Paper,* pp. 1–17.

Meskys, E. et al. (2020) 'Regulating Deep Fakes: Legal and Ethical Considerations', *Journal of Intellectual Property Law & Practice*, 15(1), pp. 24–31.

Mill, J. S. (1859) *On liberty*. London: John W. Parker and Son, West Strand.

Obar, J.A., Wildman, S. (2015) 'Social media definition and the governance challenge: An introduction to the special issue', *Telecommunications policy,* 39(9), pp. 745–750.

Poell, T., van Dijck, J. (2015) 'Social media and activist communication' in Atton, C. (ed.) *The Routledge Companion to Alternative and Community Media.* 1st London: Routledge, pp. 527–537.

Pollicino, O., Somaini, L. (2020) 'Online disinformation and freedom of expression in the democratic context: The European and Italian responses' in Baume, S. et al. (eds.) *Misinformation in referenda.* 1st London and New York: Routledge, pp. 171–193.

QUINTAIS, J. P. (2020) 'The new copyright in the Digital Single Market Directive : a critical look', *European Intellectual Property Review,* 42(1), pp. 28–41.

TUCKER, J. A. ET AL. (2018) 'Social media, political polarization and political disinformation: a review of the scientific literature', Hewlett Foundation, pp. 1–95. Available at: https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf (Accessed: 3 May 2021).

SARAPIN, S., MORRIS, P. (2014) 'When "Like"-Minded People Click: Facebook Interaction Conventions, the Meaning of "Speech" Online, and Bland v. Roberts', *First Amendment Studies,* 48(2), pp. 131–157, https://doi.org/10.1080/21689725.2014.962557.

SCHMITZ-BERNDT, S., BERNDT, C. (2018) 'The German Act on Improving Law Enforcement on Social NEtworks (NetzDG): A Blunt Sword?', *University of Luxembourg Working Paper,* pp. 1-41. Available at: https://orbilu.uni.lu/handle/10993/45125 (Accessed: 7 May 2021).

SELTZER, W. (2010) 'Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment', *Harvard Journal of Law & Technology,* 24(1), pp. 171–232.

VERSTRAETE, M. ET AL. (2017) 'Identifying and Countering Fake News', *Arizona Legal Studies Discussion Paper,* 17(15), pp. 1–38.

VOLOKH, E., FALK, D. M. (2012) 'Google First Amendment Protection for Search Engine Search Results', *Journal of Law, Economics & Policy*, 8(4), pp. 883–899.

ZUCKERBERG, M. 'A Blueprint for Content Governance and Enforcement', 15 November 2018, https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634

Chapter X

The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism

Marcin Wielec, Bartłomiej Oręziak, Aleš Rozehnal, Davor Derenčinović, Dušan V. Popović, Gábor Hulkó, András Koltay, Kristina Čufar, Sanja Savčić

1. Introduction

This chapter summarizes the considerations made to date in the previous chapters of this scientific monograph.

2. The impact of digital platforms and social media on the freedom of expression and pluralism

The chapter by Marcin Wielec investigates the impact of digital platforms and social media on freedom of expression and pluralism. The author attempts to determine the scale of the influence, benefits, and dangers of the existing operating structure of digital platforms and social media.

Marcin Wielec, Bartłomiej Oręziak, Aleš Rozehnal, Davor Derenčinović, Dušan V. Popović, Gábor Hulkó, András Koltay, Kristina Čufar, Sanja Savčić (2021) The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism. In: Marcin Wielec (ed.) *The Impact of Digital Platforms and Social Media on the Freedom of Expression and Pluralism,* pp. 311–322. Budapest–Miskolc, Ferenc Mádl Institute of Comparative Law–Central European Academic Publishing.

https://doi.org/10.54237/profnet.2021.mwsm_10

The author discusses the most important regulations in Poland regarding freedom of expression, censorship, and fake news. These include the Constitution of the Republic of Poland of 1997, the Act of September 15, 2000, Code of Commercial Companies, the Polish Criminal Code, the Polish Civil Code, the Act of December 18, 1998, on the Institute of National Remembrance – Commission for the Prosecution of Crimes against the Polish Nation. The author also explores a draft act on the protection of the freedom of social network users, which was sent to the chancellery of the Prime Minister on January 22, 2021, with a request for entry in the list of legislative work of the Council of Ministers.

The author reviewed national legislation for the admission of digital platforms and social media to individual country markets (organizational form, country branch office, legal obligations, operating restrictions, etc.) and analyzed the legal liability of the creation, dissemination, and use of fake news from the perspective of administrative, criminal, and civil laws.

3. Censorship on digital platforms and social media versus freedom of expression and pluralism in the Republic of Poland

Bartłomiej Oręziak's chapter raises an important issue from the current social perspective in the Republic of Poland and considers freedom of expression and pluralism in opposition to the issue of censorship in digital platforms and social media. This chapter provides a scientific analysis of censorship on digital platforms and social media in relation to freedom of speech and pluralism from the perspective of the Republic of Poland. This analysis comprises three main research segments that constitute the basic axis through which the problem is considered.

According to the author, censorship should be divided into censorship permitted by law and censorship not regulated by law. The author considers it necessary to discuss censorship first, which remains lawful and concerns content posted on the Polish Internet, based on the European Union (EU) regulations regarding the principles of copyright on the Internet as defined in the latest case law of the Court of Justice of the EU. The author later refers to research issues concerning censorship in relation to digital platforms and social media in Poland.

This chapter recognizes the fundamental importance of the standard of freedom of expression and pluralism established in Poland, as well as the normative grounds for any restrictions in this regard. In addition, the author analyzes the compliance of the Polish legal system to the regulation of censorship with the standards of human rights protection.

4. The regulation of social media platforms in Hungary

The chapter by András Koltay, devoted to the regulation of social media platforms in Hungary, presents an overview of various issues related to the functioning of social media platforms that have a significant impact on society. This chapter addresses legal regulations and case laws in the Hungarian legal system. The first part covers general issues related to the definition of censorship and its application, as well as various issues regarding the interpretation of censorship in relation to social media and its various manifestations. The second part focuses on the legal measures available to combat fake news and disinformation.

This chapter states that the legal relationship between social media platforms and their users (which is not affected by the constitutional doctrines of freedom of speech) is also regulated by law through an agreement concluded by and between the parties. The author draws attention to the issues of law enforcement in relation to social platforms. However, the author finds that it is not possible to enforce the principles and doctrines of freedom of speech in the online world with as much fervor as can be employed in offline spaces. With the advent of the Internet, the right to freedom of speech has entered a new phase of development with unforeseeable consequences.

The author asserts that government decision-makers and public policymakers need to adopt a systemic approach that considers the distinctive features of gatekeepers' activities, tracks their changes, provides an accurate definition of what is expected from gatekeepers and what they might expect from the law, and accurately establishes the duties and scope of liability of gatekeepers. The impact of gatekeepers on public communication and the strengthening of private regulations necessitate the use of new, creative, and innovative regulatory methods and institutions, the invention of new methods of establishing rules, and the degree of cooperation between public and private actors, which is unprecedented in this field.

As the author states, regarding the problem of fake news, the current doctrine of freedom of speech as applied in Europe does not exclude the prohibition of publication of falsehoods; hence, these cannot enjoy general constitutional protection. False statements of fact can, in certain cases, be restricted. However, the general prohibition of false statements is difficult to imagine. Simultaneously, this is a serious and massive problem for public communication and discussion of public affairs, especially on large online platforms. Any possible regulation is either contrary to the principles of freedom of speech or is likely to be ineffective. For the time being, states seem to accept their inability to regulate the public sphere without the platforms and deliberately hand over their former exclusive state function of setting the boundaries of freedom of speech to the platforms.

5. The impact of digital platforms and social media on the freedom of expression and pluralism in Serbia

The chapter by Sanja Radovanović addresses the issue of freedom of speech in relation to Internet activities. The author points out that the Constitution Act of Serbia protects freedom of expression as a fundamental right but simultaneously constitutes a framework for its restriction. In that sense, freedom of expression could be limited by the rights and reputation of others to uphold the authority and objectivity of the court and protect the public health, democratic social morals, and national security of the Republic of Serbia. To fulfill both aims, the provisions of several laws prescribe certain content that could result in violation or jeopardizing others' rights. These were not numerus clausus cases. In that sense, if some content is not explicitly recognized as inappropriate, this does not mean that it is allowed. The court, deciding on an eventual dispute, determines whether expressed content is convenient to violate someone's right or which one of the confronted rights needs protection in a specific situation.

The provisions analyzed in the chapter on Serbian law concern the media under Serbian jurisdiction. Regarding media service providers out of the jurisdiction of Serbian law and authorities, and in particular social networks that are unregulated by state law, it appears impossible to control content distributed among users and to the public. The most common problem that emerges on a daily basis occurs when the Internet provider enables the posting of illegal content and sharing among users. The liability of Internet intermediaries—information society service providers in the law of the Republic of Serbia—is normally regulated by the Law on Electronic Commerce.

The chapter notes that traditional media in Serbia still take the lead in shaping public opinion and social trends. However, the increasingly frequent reactions of traditional media to events on social networks indicate the gradual influence of social networks and their inevitable inclusion in media flows. Nevertheless, social networks do not (at least directly) fall under Serbian law and jurisdiction.

The direct enforcement of domestic law on the Internet is possible only within the field in which the state has sovereignty, which is expressed through territorial and personal authority over certain segments of architecture and content. The only way for a state to fully implement its legal system on the Internet is to take full control of the physical and logical layers of the system.

The social harm incurred by spreading false news in Serbia has been recognized by the Criminal Code—concretely by the offense entitled Causing panic and disorder. Based on Serbian case law, the more frequently invoked claim against fake news creators seems to be for monetary damages. However, unlike the specific rules on damage compensation in cases in which harmful content is provided by journalists, on the same request against the media due to damages caused by fake news, the general rules of tort law are applicable. This is due to the law of

public information and media, which stipulates the liability of the journalist and editor-in-chief for damages. Previous observations considered Serbian laws that could be applicable in the case of fake news when it caused relevant legal consequences. From a legal perspective, it would be considerably difficult to define who should be the arbiter of truth.

6. Social media, freedom of expression, and the legal regulation of fake news in Croatia

The chapter by Davor Derenčinović focuses on freedom of expression on social networks and its limitations. Special consideration is given to censorship on the Internet and the responsibility for disseminating fake news in cyberspace. Analyzing the advantages and disadvantages of the legal regulation of social networks, the current and prospective Croatian legal framework is presented and compared to the relevant legislation adopted recently in some European countries. Particular emphasis is given to the issue of service providers' responsibility for content generated by users, which was subject to deliberations of the European Court of Human Rights in several important cases. In summary, this paper aims to scrutinize and analyze regulations and procedures, identify their weak points, and offer proposals to improve dysfunctional legislation and ineffective implementation of Internet and social network policies.

According to some estimates, in 2019, slightly more than 50% of people worldwide had access to the Internet, while in 2009, this number was significantly lower (less than 5%). Over a year later, that number was estimated to be over five billion people, around 65% of the global population. According to some estimates, in 2017, there were about 2.86 billion social media users, with 3.6 billion in 2020. It is estimated that by 2025, about 4.41 billion people will have profiles on social networks.

This exponential growth of users has not been coupled with increased media literacy, knowledge about the risks of victimization on global networks, or public awareness of harmful content. Undoubtedly, this social context has created confusion and disorientation among people, most of whom use the Internet and social networks, making them more vulnerable to victimization, abuse, and manipulation both physically and in cyberspace.

Abuse in cyberspace is characterized by the phenomenon of discrimination between social media and networks. There are numerous examples of the use of speech that does not enjoy protection under Art. 10 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. The jurisprudence of domestic courts and the European Court of Human rights confirm elements

of a legitimate aim and proportionality principle in restricting the speech that, without any function in a democratic society, causes harm to others.

This chapter also focuses on preventing discrimination in cyberspace and other forms of expression that threaten European values, democracy, human rights, and the rule of law, such as disinformation campaigns aimed at disseminating fake news on social media and networks. In this context, the author focuses on the concept of responsibility of electronic media for user-generated content and prevention-based models aimed at raising public awareness about harmful Internet content and increasing media literacy.

7. Legal aspects of content censorship on social networks in Slovenia

The chapter by Kristina Čufar addresses the regulation of hate speech on Facebook and the problematic spread of misinformation and disinformation on social media. It approaches these issues by untangling the complex network of private and public regulation, paying special attention to Slovenian legislation, case law, and scholarship on the subject. The topics of hate speech and fake news are discussed, and Facebook is chosen as an example because it has the most users and the most diverse user structure of all social networks popular in Slovenia. Private censorship is a controversial topic. However, it must be acknowledged that social networks must engage in content moderation to provide a safe space for their users, remove illegal content, and appease the public, existing and potential users, states, partners, and advertisers. Therefore, the central question is not how to stop social networks from moderating content but how such practices are regulated and should be regulated in the future.

In the current era, Internet spaces are dominated by profit-driven enterprises, the algorithms of which sometimes contribute to the rise of incendiary speech and misinformation and disinformation. More speech does not necessarily mean better speech—hate speech, threats, and insults are often used to silence certain groups and may harden the pluralism of public debates. These trends are emerging in Slovenia, as social media allows anyone to express and circulate their ideas, and vulgar and offensive language often trumps nuanced discussions. People thus find themselves targeted and silenced by anonymous users, which contributes to the polarization of society and places freedom of expression up for grabs, available to the loudest and most aggressive speakers. The idea of democratic debate, conversely, presupposes a minimal level of civility and the use of arguments. As a young democracy, Slovenia has cultivated a permissive attitude toward freedom of expression, owing to the abuse of the prohibition of hostile propaganda in the former Socialist Federal Republic of Yugoslavia. Nevertheless, the Slovenian legal

system restricts individuals' freedom of expression with the freedoms and rights of others to ensure open participation in democratic debate and prevent the abuse of this freedom. Case law involving problematic expressions on social networks reveals that social networks have certain particularities compared to other forums of expression.

The chapter situates the existing Slovenian regulations in the EU legal framework. It considers the regulative approaches of social media companies on Facebook as an example to demonstrate the complexity of private content moderation and its intersections with national and transnational regulation. This chapter presents Facebook's rules and procedures for content moderation and engages with the phenomenon of fake news and its perception, unpacked upon a classification of different types of misinformation and disinformation common in the Slovenian (social) media-sphere. As social media platforms host user-generated content and do not create or edit the content on their platforms, they are not considered to be media platforms under Slovenian legislation. As host service providers, they are excluded from the liability for the content they host if they are not aware of its illegality and must only remove it once the illegality of the content is officially established. While users cannot demand a reinstatement of a deleted post under Slovenian legislation, they may demand the legal removal of an illegal post.

The ability to create and disseminate misinformation and disinformation is not systematically regulated in Slovenia; however, such activities may result in legal accountability. While misinformation and disinformation on social media influence people's perceptions of reality, it is extremely difficult to efficiently regulate it. It is impossible to draw a fine line between necessary regulations and freedom of expression or between opinions and lies. Similarly, the fine line between censorship that stifles public debate and necessary content moderation is a political question that can never be conclusively answered. The regulation of expression on social networks and related dilemmas are not unique or limited to Slovenia and will remain to be governed by social networks, states, transnational entities, and other non-state actors. States lack the infrastructure to govern expressions on social networks effectively. Nevertheless, blindly trusting social network companies to discriminate adequately between legal and illegal expression is insufficient. Well-meaning obligations imposed on social networks to remove illegal content may result in the excessive removal of user-generated content; therefore, the cooperation and inclusion of a wide scope of actors are of utmost importance. More transparency, democratic oversight, and redress procedures are thus necessary. Furthermore, users should be adequately informed about social network companies' modus operandi (the algorithmic architecture of their platforms, use of data, advertising practices, options to opt out, rules and procedures for content moderation available in local language). They should have a say in the rules and procedures and be recompensed for their data and time. The EU has an important role to play in this process and is attempting to address the described issues and

curb the power of technological companies with the proposed Digital Services Act package, the final shape and effects of which remain to be seen.

9. The role of social media in shaping society

The chapter by Aleš Rozehnal addresses social media's role in shaping society. The author indicates that most citizens, and therefore voters, use social media as their primary source of information and news. The Internet and social media thus shape our democratic dialog. Social media cannot stand above the law, but users should have the right to publish what they want, even if they risk consecutive sanctions by crossing the established limits.

In the Czech intellectual environment, the news presented by social media is what is recognized as the truth. If social media states that an event is true, it will be established as the truth regardless of the reality. The freedom of speech and the right to obtain information thus become imaginary because the only space for public discourse is in social media. Social media establishes the topics of discourse as well as the arguments and participants thereof. As opposed to autocratic censorship, democratic censorship is no longer based on omitting and deleting data but on the gathering, saturation, and surfeit of information. Thus, information is now distorted by volume.

The censorship today looks different and has different intentions in comparison with the past. It is based on more complex financial and commercial criteria, contrary to authoritative censorship. The flooding of information masks the lack of relevant information, and the images are often false and conceal the reality.

Everyone has the right to publish whatever they consider appropriate; this is the essence of freedom of speech. However, if they publish anything against the law, they must take responsibility for their actions. Freedom of speech is one of the basic pillars of a democratic society and one of the main conditions for self-development and self-fulfillment. This applies not only to information and opinions that are well-received and judged as non-aggressive or neutral but also to those that are aggressive, shocking, or irritating.

Freedom of speech and expression includes the free market of ideas in which false, criminal, and harmful doctrines will be overcome by true statements and right opinions. This freedom of the free market of ideas and the defeat of false ideas is beneficial to society as a whole. However, the issue is complex as freedom of speech has not only outer limits—limits by legal regulations—but also inner limits—immanent to this freedom. This is because freedom of speech also includes responsibility for the speech, which does not mean moral or philosophical responsibility but legal responsibility.

Internet use has brought a new dimension to the expression of freedom of speech. The easy dissemination of information in cyberspace is such a change in the amount of information that it has resulted in a change in its quality and thus in the understanding of freedom of speech. Anyone can speak to numerous people with no physical or mental effort while receiving immediate responses. Moreover, this is fully or partially anonymous and contrary to the publication of articles in standard media that are highly elitist and extremely plebeian. At first sight, free and universal access to the Internet appears to have enabled as many people as possible to establish their own ideas in the social consciousness. For the first time in history, everyone has the same opportunity to accept others' and disseminate their own perspectives and thus participate in free civil society life. Thus, the Internet is a highly democratizing media environment that could strengthen human rights and civil liberties. Instead, we witness a mass of hate speech on the Internet, mainly surrounding discussions related to published issues, especially on news and journalistic servers. This hate speech is usually of such a characteristic that it interferes with the personal rights of other people in discussions or people discussed in the main issue, or it is so rude and vulgar that it violates the basic rules for civil coexistence. Such speech is often racist and xenophobic, proclaiming intolerance and contempt for democratic systems and other people's rights.

10. The impact of digital platforms and social media on the freedom of expression and pluralism in Slovakia

The chapter by Gábor Hulkó presents the regulation of social media platforms in connection to freedom of speech, content censorship, and "fake news" in Slovakia. The author highlights the relevance of social networks in Slovakia and discusses various aspects of the liability of media platforms and users. In general terms, two main questions arise concerning the state regulation of media platforms: one is the assessment of disputes and legal liability between users, and the other is the issue of the legal liability of platforms. In the Slovak legal system, users can sue each other in the usual way of the offline world, or they can conventionally accuse if they suspect that a crime has been committed. Legal procedures are the same in the online world. The regulation of the liability regime for content on social media platforms is a distinct matter, as it tampers with different questions: first, the responsibility of social media platforms for user-uploaded content; second, the reaction of social platforms to this uploaded content, such as banning users' posts and deleting (censoring) information. In this regard, social media platforms can influence the flow of information locally and globally; thus, they *de facto* intervene with individuals' freedom of expression and right to information.

The freedom of expression and the right to information are guaranteed on a constitutional level in the general regulatory framework in Slovakia. Furthermore, censorship is prohibited on a constitutional basis. However, according to the prevailing doctrine, the concept of "censorship" is only relevant in cases in which the state faces the individual, meaning that censorship is only regarded as the actions of state organs. Therefore, this constitutional rule does not apply to the actions of private individuals or corporations capable of limiting, banning, or *de facto* censoring the views or opinions of others.

As pointed out in the chapter, social media platforms are not specifically regulated in the Slovak legal system, as subsidiaries rule the regulation for transposing the rules of the EU e-commerce Directive into national law can be applied. Regarding the liability of e-service providers and social media platforms, Slovakia follows Delfi AS v. Estonia's verdict.

The question of deleting user-uploaded information can be considered a bit of a gray area, as this type of information removal in *stricto sensu* does not qualify as an act of censorship. Only the interventions of the state are considered as censorship, but there is no actual legal definition for this term. Most authors define relevant conceptual features of censorship as its public power nature, meaning that the prohibition of censorship is addressed exclusively as the responsibility of the state. Interference with freedom of expression by private individuals—while not necessarily less threatening than interference by public authorities—cannot be considered as censorship under current Slovak regulations. However, some authors highlight that the concept of "censorship addressed exclusively to the state" is outdated and should be revised, as the forms of communication have changed and developed drastically. The system of public liability for content control by social media platforms is not known in the Slovak legal system with regard to alleged or actual censorship. The relationship between social media and the user is interpreted by Slovak law as a private law contract within the framework of which the user consents to the service provider to remove certain content. Therefore, if the information is deleted by the service provider, it can be challenged in court.

The concept of state intervention against fake news and similar types of disinformation is a current issue. It fully realizes that the important role of the state and its competent components is to create a mechanism to eliminate the impact of disinformation campaigns, especially through the effective identification of manipulative content and strategic communication. Therefore, a novel cybersecurity act and a governmental administrative action plan are prepared. The latter constitutes actions against disinformation as administrative tasks, for which most organs of the central state administration have tasks and obligations.

11. Freedom of expression on social networks: The international perspective

The chapter by Dušan V. Popović discusses the issue of freedom of expression in relation to social media platforms from an international perspective. It discusses the legislation of the United States of America and the regulations of the Council of Europe and the EU.

The author points out that international and national legal documents do not use uniform terminology to designate the right to participate in public debates. The First Amendment of the United States Constitution, adopted in 1791, employs the term "freedom of speech." More recently adopted legal documents, however, employ the term "freedom of expression." For example, the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR), adopted in 1948 and 1966, respectively, state that individuals have the right to freedom of expression. This right includes the freedom to seek, receive, and impart information and ideas of all kinds. The European Convention on Human Rights also employs the term "freedom of expression." The concept of "freedom of speech" has been interpreted extensively to include not only direct speech (words) but also symbolic speech (actions). The Internet has undoubtedly introduced new forms of communication—new forms of expression of opinions. For example, a "like" on a social network is a form of speech, as it represents a statement made by an Internet user.

Since their inception, social networks such as Facebook and Twitter have been legally considered private spaces. However, in recent years, social networks have been increasingly perceived as forums of public communication. In line with this tendency, the US courts have examined whether public forum doctrine can be applied to social networks.

In recent years, concerns about the societal consequences of the online dissemination of disinformation and propaganda have become widespread. New digital tools that allow anyone to spread political information easily to numerous Internet users can lead to a more pluralistic public debate and also provide a platform for extremist voices and actors seeking to manipulate the political agenda for their political or financial interests. The problem of "fake news" attracted substantial attention during the 2016 US presidential elections. Given its complexity and the different ways in which it is perceived, the term "fake news" has been less employed in legal doctrine and legal documents in recent years than the term "disinformation." This is particularly the case in the EU within the context of recent European initiatives.

The self-regulatory approach to combating fake news, which is preferred by social networks, as well as the co-regulatory approach favored by the EU, typically faces several challenges. First, conflicts of interest may occur between the social networks' need to keep users engaged and monetize their engagement and public

authorities' need to safeguard the integrity of democratic processes. Second, an enormous amount of content must be monitored, which necessarily implies the algorithmic screening of content and, consequently, the errors that could occur in that process. Third, the efficiency of fact-checking mechanisms is limited, as algorithms cannot be relied on for controlling the extreme amount of online content. Further, state-imposed direct regulation, preferred by certain European and non-European countries, focuses on illegal content while ignoring many other variants of disinformation.

Other titles in the book series
Studies of the Central European Professors' Network

Zoltán J. Tóth (ed.): *Constitutional Reasoning and Constitutional Interpretation*

Paweł Sobczyk (ed.): *Religious Symbols in the Public Sphere*

Tímea Barzó, Barnabás Lenkovics (eds.): *Family Protection From a Legal Perspective*

Design, layout
IDEA PLUS (Elemér Könczey, Botond Fazakas)
Kolozsvár / Cluj-Napoca (Romania)

Printed and bound by
AK NYOMDA
Martonvásár (Hungary)