

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 networks¹ 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.

relate to others.³ The possibility to express oneself and connect with others may contribute to the empowerment of traditionally sidelined groups⁴ 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 harassment⁶ 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; Prisljan, 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 trends²⁰ 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 Grimmelman, 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 campaigns³⁰) and civil society organizations³¹ 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 world³⁷ 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 Service⁴¹ 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 bylaws⁶² are similar to traditional corporate and non-profit 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 languages⁷⁰ 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 Boštich, 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). Distributors 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 Stajniko, 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 offa*, 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 transition¹⁶³ 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* (NetzDG) 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 ogrožitveno 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.rtv slo.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.rtv slo.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.rtv slo.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 tožilcev 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.rtv slo.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=2ahUKEwjs3cirq6vvAhXkwaIHHYjhAR0QFjAAegQIARAD&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 boleznih 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.rtv slo.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) *Pravljicnih 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 Šalomon, 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. *Za govor 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.rtv slo.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-zastraevanje-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-older/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/combating-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', *Slovensščina 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.cciagnet.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-twitthers-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 vsečki 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 vrednote*. 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., FILIPIČIČ, K., ZDOLŠEK, S. (eds.) (2018) *Veliki znanstveni komentar posebne 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 spodrsjlaju izbrisala svoj anonimni profil na Twitterju* [Online]. Available at: <https://www.rtvsl.si/slovenija/poslanka-sds-a-alenka-jeraj-bo-po-spodrsjlaju-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)
- MILOSAVLJEVIČ, 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)
- MILOSAVLJEVIČ, 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., ARGAS 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 omrežij 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-regulation'* [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., KUCHAR, 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 pravo 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 vstajništvo', *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 algoritimizacije 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: Opredelevitev 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/poraba-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 BOŠTIČ, 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-profilu-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) Official 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