Capstone Thesis

Supervisor: Prof. Mathias Möschel

Department of Legal Studies

Central European University

Academic year 2020/2021

The constitutional risks of fighting disinformation by law in Europe

By

Lisa Maria Weinberger

Email: weinberger_lisa@student.ceu.edu

Vienna, Austria

16 June 2021

Abstract

With the aim to effectively tackle the spread of false and hateful content online, some countries have introduced legislation to criminalize disinformation online. Such disinformation laws are of great concern due to their risk of deterring effects on free speech and liberal democracy. The first aim of this thesis is to provide an attempt at classifying how EU countries regulate this area. Based on the public or private nature of the decision-making body, a distinction will be made between platform regulatory laws and state regulatory laws. Secondly, this thesis will look at the key constitutional aspects of both models evolving around the risk of abuse of power, starting with the most discussed and scrutinized concern of freedom of expression to less obvious aspects such as legal certainty, separation of powers, independent oversight and criminalization of disinformation. It suggests that the loopholes for abuse of power start with the lack of legal certainty regarding the concept of disinformation, which gives the respective decision-making body great discretion when determining the illegality of content, often without judicial supervision. This problem applies to both models. Distinct safeguards present in some laws do not remedy this initial default.

In relation to the platform regulatory model, the analysis will focus on Germany as the birthplace of such laws in Europe and France, where the "Avia Law" was declared unconstitutional, as well as Austria, the most recent example in the EU. This model will be contrasted with current examples of state-based regulations, namely another French law and Hungary's COVID-19 related emergency provision. Given the comparative framework, the findings of this thesis may allow for a better overview of the developments and different models of disinformation laws in Europe and contribute to a broader understanding of the constitutional risks such laws pose given the risk of abuse of power and its implications for liberal democracy. The issues identified at a national level may also be of relevance for the EU when elaborating its own solution on how to best regulate the digital single market.

Table of Contents

Introduction

With the emergence of the internet, there was the hope that the digital space would create opportunities for a more democratic and diverse marketplace of ideas. In reality, it often facilitates spaces of divisiveness, fragmentation and segmentation. Due to the non-human mass distribution of ideas, "certain views held by real people cannot even emerge in a cacophony of false information. Certain viewpoints may therefore simply be outnumbered and vanish."[1] Undoubtedly, there is a pressing need to address these systemic issues that threaten democratic processes and prevent citizens to fully inform themselves and allow for equal access of all voices in public debates. Information is a precious commodity in a democracy and as such it must be duly protected. However, there is a tendency among governments to look for quick fixes, introducing restrictive and short-sighed solutions that are overly inclusive and restrict legitimate along with illegitimate speech.[2] Governments around the world have reacted very differently to the phenomenon but a considerable number has introduced so-called "disinformation laws", which criminalize the spread of false information. By 2019, around 40 countries introduced such laws.[3] In Europe, France and Germany received particular attention for each introducing a particular model of disinformation laws. Their structural differences will be analysed in detail in this thesis.

One of the key constitutional concerns that will be covered in this thesis is the lack of a clear and universal definition of disinformation that countries can rely on when designing such legislation. This leads to vague or ambiguous terms such as "false information"[4] or the claim

[1] Judit Bayer and others, 'Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States' (Disinformation and propaganda) (European Parliament 2019) 64

[2] Media Legal Defence Initiative, Module 3: Criminalisation of online speech (2020) 2 https://www.mediadefence.org/ereader/wp-content/uploads/sites/2/2020/06/Module-3-Criminalisation-of-online-speech.pdf accessed 13 June 2021

[3] Daniel Funke and Daniela Flamini, 'A guide to anti-misinformation actions around the world' (*Poynter*) https://www.poynter.org/ifcn/anti-misinformation-actions/ accessed 13 June 2021

[4] *loi* n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (French Law n° 2018-1202 against the Manipulation of Information)

or spread of "falsehood" or "a distorted truth"[5] being used as the basis for highly restrictive disinformation laws. The European Union (hereafter EU) makes a distinction between disinformation and misinformation whereby the latter is "false information, or dissemination of such information not necessarily in the knowledge that it is false" while disinformation is "deliberate misinformation".[6] Similarly, the European Commission's High-Level Expert Group on Fake News and Online Disinformation (HLEG) defines disinformation as "all forms of false, inaccurate, or misleading information designed, presented and promoted to *intentionally* cause public harm or for profit" [emphasis added].[7] Despite this clear distinction based on intention, EU institutions such as the European Parliament often use the terms interchangeably.[8]

While the constitutional concerns of the German and French laws have been raised by various national and international experts, the most recent statutes in Austria and Hungary have not been closely analyzed and, more importantly, have not been assessed in a comparative manner based on a constitutional framework by using separate criteria of constitutionality rather than focusing exclusively on free speech concerns. The most extensive studies on disinformation in Europe from recent years, an EU study on Disinformation and propaganda from 2019[9] as well as a Handbook on Hate Speech and Disinformation from 2020[10], do not use a comparative methodology but remain descriptive and, moreover, do not cover any laws introduced in 2020 (namely Austria, Hungary and the French "Avia Law"). Additionally, a categorization of different types of disinformation laws in Europe has not been conducted yet. Here, a distinction will be made between platform regulatory laws, where platforms themselves are required to control the content on their websites, and state regulatory laws, where content-moderation is

[5] Section 337 of the Hungarian Bill T/9790
[6] iate, https://iate.europa.eu/entry/result/3576921 accessed 13 June 2021
[7] High level Group on fake news and online disinformation, A multi-dimensional approach to disinformation (European Commission 2018) 3
[8] Judit Bayer and others, 'Disinformation and propaganda' (European Parliament 2019) 25
[9] ibid.
[10] Giovanni Pitruzzella and Oreste Pollicino, *Disinformation and Hate Speech : A European Constitutional Perspective* (Bocconi University Press 2020)

not outsourced but done by the state.

Therefore, the aim of this thesis is to provide an analytical and comparative overview of the risks of these two models of disinformation laws from a European constitutional perspective. The first part of the thesis will consist of an attempt to classify these two models of disinformation laws while the second part will focus on the crucial constitutional aspects of both models evolving around the risk of abuse of power, starting with the most discussed and scrutinized concern of freedom of expression to less obvious aspects such as legal certainty, separation of powers, independent oversight and criminalization of disinformation. In relation to the model of platform regulatory laws, the analysis will focus on Germany as the birthplace of such laws in Europe and France, where the "Avia Law" was declared unconstitutional, as well as Austria, which is the most recent example in the EU. Regarding the state regulatory model, a comparison will be made between the well-established and extensive law in France and Hungary, which is the only EU country introducing a COVID-19 related disinformation provision during the pandemic. Given the comparative framework, the findings of this thesis assist to allow for a better overview of the developments and different models of disinformation laws in Europe and contribute to a broader understanding of the constitutional risk such laws pose given the risk of abuse of power and its implications for liberal democracy.

1. Two models of disinformation laws in Europe

Since 2000, the EU has relied on the e-Commerce Directive, which aims to harmonize cross-border online services within the EU and to provide legal certainty for businesses and consumers. In relation to online service providers such as electronic communications services or social networks, online marketplaces and other hosting service providers, the Directive grants relative immunity to providers for illegal third-party content as long as they act "expeditiously" to remove or disable access to information when becoming aware of illegal activities on their

services.[11] Furthermore, the Directive encourages Member States to set up codes of conduct and prohibits to oblige intermediary providers to general monitoring, thus striving for a self-regulatory framework of intermediary providers rather than regulatory laws.[12]

However, since the e-Commerce Directive came into force, the digital marketplace has changed rapidly. The emergence of social media and its impact on political life, particularly its power to influence elections, accelerated concerns among various European governments to rely exclusively on the existing EU framework. In Germany, the *Willkommenspolitik* of Chancellor Angela Merkel in 2015 to accept over a million refugees in Germany led to a significant increase in hate speech and disinformation on social media platforms. This triggered the introduction of the German Network Enforcement Act (*Netzdurchsetzungs-Gesetz*, hereafter *NetzDG*).[13] The NetzDG is considered the European prototype when discussing contemporary disinformation laws. The German model served not only as an inspiration abroad but strikingly similar versions exist in other EU countries. Similarly, serious concerns in France of attempts to influence elections emerged during the presidential elections of 2017, which triggered the introduction of the Law against the Manipulation of Information in 2018.[14] This was followed by another French law "On Countering Online Hatred", which was introduced in May 2020.[15] This law is commonly referred to as the "Avia Law" due to Parliamentarian Laetitia Avia who proposed the law. Immediately after its adoption, it was brought before the *Conseil Constitutionnel* (hereafter CC) by a number of senators due to constitutional concerns and was

[11] Council Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) [2000] L 178/1, art 14 (1) (b)

[12] Directive on electronic commerce, art 15-16.

[13] 'Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken' (Introduction to the Draft of the Federal Government of Germany regarding the NetzDG) (*Ministry of Justice and Consumer Protection)* https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RegE_NetzDG.pdf?__blob=publicationFile&v=2 accessed 13 June 2021

[14] French Law n° 2018-1202 against the Manipulation of Information

[15] Proposition de loi, adoptée par l'Assemblée nationale dans les conditions prévues à l'article 45 alinéa 4 de la Constitution visant à lutter contre les contenus haineux sur internet le 13 mai 2020, T.A. n° 419 (Avia Law n° 419)

declared in most parts as unconstitutional in June 2020.[16] Nevertheless, as the law's structure and the subsequent decision by the CC provide valuable insights for the following assessment of constitutionality, it will be included in this analysis. In reference to the German model, Austria adopted the communications platform law (*Kommunikationsplattformen-Gesetz,* hereafter *KoPl-G*) in January 2021 while pointing out that the draft included considerations and "lessons-learned from France".[17] Since the global health pandemic due to COVID-19, more countries have followed suit. Among them in Europe are Hungary and Romania. As part of declaring a state of emergency, they announced their fight against the spread of "false news" in relation to the pandemic as well.[18] Since Romania only included the regulation as part of a temporary Presidential Decree and this thesis focuses on legislative measures, Romania will only be mentioned to highlight certain features of the Hungarian law.

This development to fight disinformation by law as outlined above is not a given as the majority of European countries refrained from criminalizing disinformation while using or amending existent media and hate speech regulations for online news.[19] Several countries, in particular Nordic and Baltic countries, took alternative action by implementing various strategies against disinformation including awareness-raising campaigns (Sweden, Netherlands),[20] digital and media literacy programs in schools (various Nordic and Baltic countries),[21] setting up task

[16] Conseil Constitutionnel, Décision n° 2020-801 DC of 18 June 2020

[17] Republic of Austria, Parliamentary correspondence Nr. 1391 of 10 December 2020 https://www.parlament.gv.at/PAKT/PR/JAHR_2020/PK1391/index.shtml accessed 13 June 2021

[18] Section 337 of the Hungarian Bill T/9790. Regarding Romania: Art 54 (4) of the Presidential Decree No. 195/2020

[19] The toolbox in this regard is already significantly greater than in countries such as the United States, where such laws would largely violate the constitutional right of free speech under the First Amendment.

[20] 'Sweden to create new authority tasked with countering disinformation' *The Local* (Stockholm, 15 January 2018) https://www.thelocal.se/20180115/sweden-to-create-new-authority-tasked-with-countering-disinformation accessed 13 June 2021
See also: Olga Robinson Alistair Coleman and Shayan Sardarizadeh, 'A Report of Anti-Disinformation Initiatives' (Oxford Technology & Elections Commission 2019) 4-6 https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/08/A-Report-of-Anti-Disinformation-Initiatives accessed 13 June 2021

[21] Emma Charlton, 'How Finland is fighting fake news – in the classrom' (*World Economic Forum*, 21 May 2019) https://www.weforum.org/agenda/2019/05/how-finland-is-fighting-fake-news-in-the-classroom/ accessed 13 June 2021

forces (United Kingdom, Denmark)[22] and expert groups (Spain)[23]. Some countries in Europe also tried but eventually failed to introduce criminal disinformation laws.[24]

As this thesis focuses primarily on successful disinformation laws, a distinction can be made between two current models of disinformation laws, which emerged in Europe over the past years. Both models focus on the dissemination of illegal third-party content online through online service providers such as social media platforms. One distinctive difference between these laws is the nature of the decision-making body and whether it is the online platform or the judiciary deciding on the illegality of the content. Based on this, a distinction will be made between platform regulatory laws and state regulatory laws. In the following, these two models will be analyzed in more detail based on their scope, applicability, notice-and-take-down mechanism, sanctions and possibilities of redress.

1.1. Platform regulatory laws

There are several key characteristics of the platform regulatory laws in Germany, Austria and the initial draft of the Avia Law in France. First, the laws' aim is very broad and does not only

[22] Peter Walker, 'New national security unit set up to tackle fake news in UK' *The Guardian* (London, 23 January 2018) https://www.theguardian.com/politics/2018/jan/23/new-national-security-unit-will-tackle-spread-of-fake-news-in-uk accessed 13 June 2021

[23] 'Spain and Russia agree to set up joint cybersecurity group' *AP News* (New York, 6 November 2018) https://apnews.com/article/00061a7eb8814fc38ebef55880c3fd18 accessed 13 June 2021

[24] An Italian version of the NetzDG was discussed in 2018 as well a draft law called "Provisions to prevent the manipulation of online information, to ensure transparency on the web and to encourage media literacy" (DDL no. 3001 – XVII Legislature "DDL Gambaro"), which would have criminalized the publication or dissemination of "false, exaggerated or biased news" online that is manifestly unfound or false. Both bills never passed parliament, but a new hate speech law (Delibera n. 157/19/CONS) was introduced in May 2019, which includes obligations for online video-sharing platforms to remove and report hateful content. Croatia also discussed the introduction of the German model in 2018 and in Ireland, a bill to criminalize the use of bots to spread political misinformation was suggested in 2017 but was not successfully pushed through due to the dissolution of parliament (Proposal Online Advertising and Social Media (Transparency) Bill 2017 (Bill 150 of 2017). Bulgaria planned to adopt a harsh criminal law to counter the spread of false information regarding the coronavirus in 2020, which was vetoed by the President.
See also: Lida Filippakis, 'Croatian anti-hate law might lead to censorship, some experts fear' *Independent Balkan News Agency* (Bourgas, 19 January 2018) https://balkaneu.com/croatian-anti-hate-law-might-lead-to-censorship-some-experts-fear/ accessed 13 June 2021
'Repressive laws, prosecutions, attacks… Europe fails to shield its journalists against the abuse of the COVID-19 crisis' (*Reporters Without Borders*, 8 April 2020) https://rsf.org/en/news/repressive-laws-prosecutions-attacks-europe-fails-shield-its-journalists-against-abuse-covid-19 accessed 13 June 2021

target disinformation but also hate speech. A clear definition is missing for both terms. Instead, reference is made to a list of existing criminal provisions which are considered "unlawful content".[25] Generally, the focus seems to lie more on hate speech than disinformation although false content can be criminalized as part of provisions such as insults, defamation or libel, holocaust denial, incitement to hatred or propaganda.

Secondly, regarding the applicability of disinformation laws, these laws refer to online service communication providers. In Germany, the *NetzDG* focuses on social networks with more than two million registered users in Germany, thus targeting big players such as Facebook, Twitter and YouTube. Social networks are defined as telemedia service providers which, for profit-making purposes, operate internet platforms that are designed to enable users to share any content with other users or to make such content available to the public.[26] This explicitly excludes platforms offering journalistic or editorial content like online newspapers and platforms for individual communication or the dissemination of specific content such as email, messaging apps or personal blogs.[27] In comparison, the Austrian law targets all online service communication providers with either more than 100.000 users per year or an annual turn-over of € 500.000.[28] E-commerce platforms that offer services or goods, non-profit-online encyclopaedias and journalistic websites are exempted.[29] In contrast to Germany, this definition

[25] In Germany, the list includes criminal insults, defamation, libel, threats, the dissemination of propaganda material or use of symbols of unconstitutional organizations, preparation of a serious violent offense endangering the state or encouragement thereof, public incitement to commit a crime, forming criminal or terrorist association, incitement to hatred, depiction of violence, defamation of religion, distribution of every kind of pornography and the falsification of relevant evidence. In Austria, the law refers in § 2 (8) to 15 provisions in total including insults, incitement to hatred, offense to religion, coercion and stalking among others. In France, Art 1 of the proposed Avia Law also made reference to various criminal law provisions and includes slander or provocation to discrimination, hatred or violence against an individual or group of individuals based on their origin, ethnicity, race, religion, gender, or gender identity, sexual orientation or disability, the glorification of criminal offences, denial or trivialization of a crime against humanity, genocide, war crime, sexual harassment and child pornography.

[26] § 1 (1) NetzDG.

[27] ibid; 'Leitlinien zur Festsetzung von Geldbußen im Bereich des Netzwerkdurchsetzungsgesetzes (NetzDG)' (*Federal Ministry of Justice and Consumer Protection,* 22 May 2018) 3 https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/NetzDG_Bu%C3%9Fgeldleitlinien.html accessed 13 June 2021

[28] § 1 (2) KoPl-G

[29] § 1 (3) KoPl-G

targets any type of online platform rather than focusing on for-profit social networks. This may raise disadvantages in competition, particularly regarding the growth of smaller companies in Austria if faced with high regulation burdens in comparison to their neighbouring countries.[30] In France, the scope of applicability initially included all online platform operators which offer an online communication service to the public or use algorithms to offer or put online content by third-parties and which meet a certain threshold of "activity" in the French territory. However, this was to be determined by decree. As a consequence, it would have been upon the discretion of the French government to decide which providers may be included by law. Generally, this could have targeted smaller platforms as well as non-for-profit platforms.

At the heart of this model is the online providers' obligation to provide for an effective and transparent complaint management procedure for unlawful content and to remove or block "manifestly illegal" content within 24 hours and any other "illegal content" within 7 days upon notice.[31] The 7-day rule was not included in the French draft. Instead, an even stricter obligation for platforms would have applied to take down content within one hour after receiving a notification by an administrative authority regarding illegal terrorist content and sexual child abuse. The fear to be flooded by take-down requests was addressed in Germany's guideline which further clarifies that a notification must be sufficiently precise to allow for a qualified examination of the potential violation against criminal law and to, consequently, qualify as a complaint.[32] If additional investigation is needed to decide on the content's unlawfulness, the deadline may be extended.[33] In the German context, the social network may also refer the

[30] epicenter.works, 'First Analysis of the Austrian Anti-Hate Speech Law (NetDG/KoPlG)' (*EDRi*, 10 September 2020) https://edri.org/our-work/first-analysis-of-the-austrian-anti-hate-speech-law-netdg-koplg/ accessed 13 June 2021

[31] Germany: § 3 (2) NetzDG; Austria: § 3 (3) 1 KoPl-G; Art 1 of the adopted version n° 419 of the Avia Law (hereafter Avia Law n° 419)

[32] 'Leitlinien (NetzDG)' (*Federal Ministry of Justice and Consumer Protection,* 22 May 2018) 7

[33] § 3 (2) No. 3a NetzDG

decision to a recognised self-regulation institution within 7 days of receiving the complaint.[34]

Furthermore, the social network must immediately notify the person submitting the complaint about the decision and its reasoning.[35] This should strengthen the victims' rights for potential civil law claims.[36] In contrast to Germany and France, which only vaguely refer to the need of an internal redress mechanism[37], the Austrian model explicitly includes a complaint and redress mechanism. This allows the user whose content was blocked to appeal the provider's decision. In Austria, any person who issues or is issued a complaint receives information about the procedure of the complaint mechanism, the decision and its basis and their right to appeal.[38] In case of an appeal, the review is primarily done by the platform within two weeks.[39] However, in case of an unsatisfactory outcome, a second-level redress can be brought in front of the arbitration body of the telecom and media regulator RTR.[40] A court is not involved in the appeal procedure to scrutinize whether the content has actually been illegal as the RTR is an extrajudicial body. In case of more than five complaint procedures in front of the RTR within a month, the supervisory authority *KommAustria* can initiate a procedure to check the functionality of the moderation procedure in place. In case *KommAustria* finds deficiencies, it may issue an administrative decision including requirements on the improvement of the moderation procedure.[41] In case of noncompliance, *KommAustria* can issue penalties of up to €10 million for companies.[42]

Furthermore, all laws oblige the respective company to provide a national representative as a contact person. In Austria, such a representative can face up to €10.000 for non-compliance and

[34] § 3 (2) No. 3b NetzDG
[35] § 3 (2) No. 5 NetzDG; § 3 (1) KoPl-G; Art 2 Avia Law n° 419
[36] Sandra Schmitz and Christian M. Berndt, 'The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword?' (A Blunt Sword) [2018] SSRN, 19
[37] Art 4 (5) Avia Law n° 419
[38] § 3 (3) 2 KoPl-G
[39] § 3 (4) KoPl-G
[40] § 8 (2) KoPl-G
[41] § 9 (1) and (2) KoPl-G
[42] § 10 (1) and (2) KoPl-G

companies up to € 58.000 for failure of providing information.[43] An infringement of the German law can lead up to €5 million for individuals and €500 million for companies for a lack of a functioning compliance system or failure of public reporting.[44] The amount of the administrative fine as well as the initiation of the regulatory fine proceeding lies at the discretion of the competent administrative authority, which is the Federal Office of Justice in the case of Germany.[45] While the *NEtzDG*'s aim is to punish "systemic failure" of decision-making practice of social networks to avoid the risk of "over-blocking" as outlined in the published explanatory notes to the Act for setting fines, this is not specified in the Act itself.[46] The assessment of whether a systemic failure exists is, however, overseen by a court.[47] In France, the maximum fine for individuals would have been € 250.000 and €1,25 million for companies in case of failure to comply with the notice-and-take-down system within the given timeframe.[48] Furthermore, the Superior Audiovisual Council (hereafter CSA) as a supervisory and regulatory organ would have had the possibility to impose an administrative fee of up to €20 million or 4% of the annual turnover for serious and recurrent failures to remove content.[49]

Lastly, service providers must produce a report on the handling of complaints about unlawful content on their platforms. In contrast to Germany and France, the report in Austria must include content which had been removed on the basis of internal community standards. This is a significant advantage in comparison to Germany, where most blockages are based on internal community guidelines and thus not reported on.[50] Additionally, under Austrian law, platforms must save the deleted content including information about the posted content for 10 weeks for

[43] § 10 (4-6) KoPl-G
[44] § 4 (2) and (4) NetzDG in conjunction with § 20 (2) 3 OWiG (Code on Administrative Offences)
[45] § 47 OWiG
[46] See an overview regarding the limitations in: Sandra Schmitz and Christian M. Berndt, 'A Blunt Sword?' [2018] SSRN, 19
[47] ibid, 25
[48] Art 6 Avia Law n° 419
[49] Art 7 II Avia Law n° 419
[50] epicenter.works, 'First Analysis of the Austrian Anti-Hate Speech Law (NetDG/KoPlG)' (*EDRi*, 10 September 2020)

evidence and redress reasons,[51] but the Act does not include an automatic obligation to forward blocked illegal content to the Austrian authorities. This is different from Germany where such an obligation has been introduced in 2020 by Parliament but has not come into effect as the amended law was not promulgated by the President due to constitutional concerns such as data privacy.[52]

Concluding, the platform regulatory model focuses mostly on online platforms as the primary responsible actor to remove or block unlawful content as defined by domestic law. This is done upon notice within a specific timeframe of 24 hours or 7 days. In case of recurrent non-compliance, regulatory bodies such as the CSA in France, the RTR in Austria or the Federal Office of Justice in Germany may impose high administrative fees on platforms ranging from €10 to €500 million.

1.2. State regulatory laws

Unlike platform regulatory laws, the model of state regulatory laws does not rely on existent provisions of the domestic criminal code but introduces an entirely new offense: the criminalization of spreading false information. Furthermore, the removal or blockage of content is not outsourced to the respective online platform, but it is the judiciary which orders the removal of content.

Regarding the scope of the laws in Hungary and France, the French Law against the Manipulation of Information is the narrower as it focuses exclusively on the spread of "false information" during election periods, which is defined as three-months before an election.[53] The Hungarian law is directly linked to the COVID-19 pandemic, where Parliament passed an

[51] § 3 (3) 3 and 4 KoPl-G
[52] Georg Mascolo and Ronen Steinke, 'Bedenken in Bellevue' *Süddeutsche Zeitung* (Munich, 17 September 2020) https://www.sueddeutsche.de/politik/hate-speech-hasskriminalitaet-gesetz-steinmeier-1.5034929 accessed 13 June 2021
[53] Art 1 (2) French Law n° 2018-1202 against the Manipulation of Information

Act "On Protection Against the Coronavirus" on 30 March 2020 which criminalizes, among other measures[54], the claim or spread of "falsehood" or "a distorted truth" in relation to the emergency with up to three years of prison if it is "suitable for alarming or agitating a large group of people at the site of the emergency" and up to five years of prison if "suitable for obstructing or preventing successful protection" of the public.[55] In comparison, the sanction for a violation of the French law is one year in prison and a fine of €75.000.[56] Such a law may apply to any individual including journalists (see in comparison that journalistic content is explicitly excluded under platform regulatory laws). While the Hungarian state of emergency law did not include a predefined end date, Parliament repealed the state of emergency in June 2020. However, due to another bill, which was approved by Parliament, the government may impose another state of emergency in case of a "state of health emergency" at any time.[57]

Romania also introduced emergency measures to fight disinformation during the COVID-19 pandemic between March and May 2020. However, as it is a decree issued by the President and not a legislative measure as such, the regulation is very different in nature in comparison to the Hungarian provision.[58] Here, it is the National Authority for Administration and Regulation of Communications (hereafter ANCOM) which was authorized to block access to any content on an electronic communications network that promotes "false news" regarding the development of COVID-19 and protection and prevention measures. Subsequently, providers of electronic communications networks were obliged "to immediately block access to that content and

[54] The second main provision gives government the power to rule by decree.
[55] Section 337 of the Hungarian Bill T/9790. An English translation of the Act is available here: https://hungarianspectrum.org/2020/03/21/translation-of-draft-law-on-protecting-against-the-coronavirus/ accessed 18 March 2021 The Hungarian version is available here: https://www.parlament.hu/irom41/09790/09790.pdf accessed 13 June 2021
[56] Art 1 French Law n° 2018-1202 against the Manipulation of Information in conjunction with Art 112 of the Electoral Code (Code electoral)
[57] BBC News, 'Coronavirus: Hungary votes to end Viktor Orban emergency powers' (16 June 2020) https://www.bbc.com/news/world-europe-53062177 accessed 13 June 2021
[58] The decree proclaimed a 30-day state of emergency in March, which was extended for another 30-days in April 2020 and ended in May 2020. See Presidential Decree No. 195/2020; Presidential Decree No. 311/2020.

inform users"[59] upon the ANCOM's request. In comparison to Hungary, the Romanian decree includes no sanctions for the non-compliance of service providers. It is rather the general lack of transparency of when and why websites are blocked by the national authority, which are of concern.[60]

A significant difference to the platform regulatory model is that in France and Hungary, the decision-making body is the judiciary, which includes the possibility for judicial review. The French law enables legal action against the circulation of false information by evoking a judicial decision to delete content within 48 hours of referral.[61] The Hungarian law does not foresee such a timeframe. Furthermore, no additional guideline is given for how the Hungarian judges should determine the falsity of information. This is different in France, where the CC issued reservations of interpretation as part of the preliminary legislative review process. The CC held that the term "false information" may only apply to "inaccurate or misleading allegations or imputations of a fact likely to alter the sincerity of the upcoming election. These allegations or imputations do not cover opinions, parodies, partial inaccuracies or mere exaggerations."[62] This means that falsity must be demonstrated objectively. Furthermore, the dissemination of such allegations or imputations must be artificial or automated, occur on a massive scale and be deliberate.[63] Additionally, the risk of influencing the election must be manifest and the judge assigned to the case must only apply the least infringing measures and meet the criteria of

[59] Art 54 (4) of the Presidential Decree No. 195/2020 http://legislatie.just.ro/Public/DetaliiDocument/223831 accessed 13 June 2021

[60] The ANCOM acts on the request of the Ministry of the Interior but it is unclear what criteria are used to evaluate the legitimacy of content. The procedure as such is not clearly defined in the Presidential Decree. The Strategic Communication Group, which is an entity under the Ministry of the Interior responsible for communication, clarified on its website that the analysis of content will be carried out on a case-by-case basis by them based on a "systematic and deliberate dissemination of false information, ignoring the call for fair and objective information". This analysis then forms the basis for the Interior Minister's decision and the subsequent request towards ANCOM to implement the decision. See also: Marcel Gascón Barberá, 'Romania's Drive to Censor 'Fake News' Worries Activists' (*Balkan Insight*, 27 April 2020) https://balkaninsight.com/2020/04/27/romanias-drive-to-censor-fake-news-worries-activists/ accessed 13 June 2021

[61] Art 1 French Law n° 2018-1202 against the Manipulation of Information

[62] ibid para 21

[63] ibid

necessary, appropriate and proportionate measures.[64] Such clarifications do not exist in the Hungarian context.

Concluding, the state regulatory laws are not as homogenous in scope and structure as the platform regulatory model. While the French law focuses on election periods and uses a timeframe of 48 hours to issue the take-down of content, the Hungarian emergency provision does not include such a fast-track procedure and, moreover, ended in May 2020. The vague formulation is of concern in both cases, which will be discussed in more detail in the following section on legal certainty. Overall, the high penalty of five years of imprisonment in Hungary is particularly striking.

2. The constitutional risks of fighting disinformation by law

Regulating false news evokes several constitutional issues which ultimately boil down to the question of whether curtailing free speech in the name of democracy risks governments to become the arbiter of the truth. This section of the thesis will analyze the key constitutional issues of both models of disinformation laws, starting with the issue that has received most attention in relation with disinformation laws, freedom of expression, and moving to less evident aspects of constitutionalism which have emerged from legislative debates and relevant case-law. The latter include legal certainty, separation of powers and the (in)dependence of the decision-making body, independent oversight and the right to appeal and the need for criminalization.

2.1. Curtailing free speech for the sake of more safety?

The right to freedom of expression is an internationally protected fundamental human right.[65]

[64] ibid para 23ff
[65] Art 19 of the Universal Declaration of Human Rights, Art 19 of the International Convent on Civil and Political

The UN Special Rapporteur (hereafter UNSR) noted that it is an "enabler of other rights", impacting economic, social and cultural rights[66] as well as political rights such as the right to vote.[67] The European Court of Human Rights (hereafter ECtHR) views freedom of expression as one of the essential foundations of a democratic society, one of the basic conditions for its progress and for the development of every person.[68] Nevertheless, freedom of speech is not an absolute right but subject to certain limitations. In the European context, any limitation must be provided by law, pursue a legitimate aim and be necessary for a legitimate purpose as outlined in Art 10 (2) of the European Convention of Human Rights (hereafter ECHR). The Court also made clear that Art 10 (2) ECHR is "applicable not only to "information" or "ideas" that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that "offend, shock or disturb the State or any sector of the population."[69] Thus, generally offensive speech is protected as well. However, as a supra-national Court, the Member States have a certain margin of appreciation in assessing whether a social pressing need exists.[70] Additionally, freedom of expression as a fundamental right is only protected with due respect to other interests and in such a case of competing interests, the broad concept of "public interest" often prevails.[71] Lastly, the ECtHR views Art 10 ECHR not only as the right to express oneself but also as the right to receive information.[72] It is the implicit functional understanding of the right to be informed that provides the basis for deriving a right to the correct information.[73]

This approach is in stark contrast to the United States where the Supreme Court grants free

Rights, Art 10 of the European Convention on Human Rights and Art 11 of the Charter of Fundamental Rights of the European Union.

[66] United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2011] A/HRC/17/27

[67] United Nations Human Rights Committee, General Comment No. 34 [2011] CCPR/C/GC/34 para 20

[68] *Handyside v United Kingdom* App no 5493/72 (ECtHR, 7 December 1976) para 49

[69] ibid

[70] ibid

[71] See András Sajó and Renata Uitz, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (Oxford University Press 2017) 413

[72] Giovanni Pitruzzella and Oreste Pollicino, *Disinformation and Hate Speech : A European Constitutional Perspective* (Bocconi University Press 2020) 12

[73] ibid, 58

speech an inherent meaning rather than viewing it as an instrumental value and a means toward some other goal.[74] Consequently, only specific categories of speech such as incitement to commit lawless conduct or defamation are excluded from the general free speech protection.[75] Mere falsity as such is not enough to push a speech outside the roam of protection under the First Amendment as it needs "more speech, not less"[76] in order to get closer to the truth.[77] The reason for such an absolute approach is the Supreme Court's concern that allowing for the criminalization of false statements would endorse government authority to compile a list of subjects about which false statements are punishable and give the government a broad censorial power to control unfavourable speech.[78] Singling out speech is simply deemed too dangerous. By contrast, on the other side of the Atlantic, very harsh restrictions on freedom of expression apply to acts related to terrorism, offence against religion and hate speech. However, with the exception of the Holocaust denial, these offenses are all context-dependent.[79] This means that they only apply in specific cases under specific circumstances. In the case of the platform regulatory model, assessing the context is now outsourced to private entities.

The platform regulatory model adds another complication as the risk of censorship shifts from the government to the online service providers. Consequently, the duty to quickly block or remove content under the threat of harsh punishment creates an incentive for overblocking. This is why intermediaries have been exempted of liability in the past in Europe (see the e-Commerce Directive) as well as in the United States (see Section 230 of the Communications Decency Act). This precautionary stance changed in the European context with the decision of *Delfi AS v. Estonia in 2015* when the ECtHR ruled in favor of imposing the obligation to intermediaries

[74] See on truth-based approach Schenk v. United States, 249 U.S. 47 (1919) and its democratic justification Whitney v. California 274 U.S. 357 (1927)
[75] Chaplinsky v. New Hampshire 315 U.S. 568 (1942)
[76] Whitney v. California 274 U.S. 357 (1927) para 44
[77] United States v. Alvarez 567 U.S. 709 (2012)
[78] ibid
[79] Judit Bayer and others, 'Disinformation and propaganda' (European Parliament 2019) 91

to take down content "without delay" after publication. The dissenting opinion in this case warned of the risk of collateral censorship: "Governments may not always be directly censoring expression, but by putting pressure and imposing liability on those who control the technological infrastructure (Internet service providers etc.), they create an environment in which collateral or private-party censorship is the inevitable result."[80] While *Delfi* effectively pushed intermediaries to monitor and filter content online, exercising prior restraint, the dissent's assessment relates to the very short 24-hour framework of the German and Austrian disinformation law as well where the set-up leads to the same dynamic of pushing the platforms to "stay on the safe side" when receiving a request to take down content. In relation to the French draft of the Avia Law, the digital rights organisation EDRi has pointed to several other negative consequences from a technical point of view: "[The 24-hour framework] strongly encourages providers to use automated moderation tools, which frequently carry with them significant risks of false positives and false negatives: it is well-established that they are not equipped to make complex judgments in cases, which are highly context dependent. They cannot tell infringement apart from legal uses like parody, counter speech, or legitimate political dissent. Consequently, legal content will inevitably be taken down."[81] The short framework was also at the heart of the decision by the CC for declaring the Avia Law in most parts unconstitutional. The CC held that these provisions "can only encourage online platform operators to remove the content reported to them, whether or not it is manifestly illegal", thus leading to an incentive to censor free speech.[82]

Restrictions on free speech should not put the right itself in jeopardy. The question is when the balance tips over and free speech restrictive laws are used (or abused) to stifle public debate on

[80] See also the Joint Dissenting Opinion of Judges Sajó and Tsotsoria in *Delfi AS v Estonia*, App no 64569/09 (ECtHR, 16 June 2015) para 2
[81] EDRi, 'Contribution to the examination of France's draft law aimed at combating hate content on the internet' (Brussels, 18 November 2019) https://edri.org/wp-content/uploads/2019/11/20191118_EDRiCommentsEC_FrenchAvialaw.pdf accessed 13 June 2021
[82] Conseil Constitutionnel, Décision n° 2020-801 DC of 18 June 2020 para 19

matters of general or specific interest. The categorical approach in the United States prevents the Supreme Court to easily lose sight of what is at stake, that is the fundamental value of free speech in a democratic society. For a judge, this is challenging to tackle as "the constitutional problem is narrowed down to a single case before the court. The consequence is that it will be easy to sacrifice many a single individual for the community. It is simply not on the horizon that the right denied in the case is also the right of many other people and it has importance for society as a whole. After all, fundamental rights are broad societal value considerations and part of the common good or the public interest."[83]

2.2. Legal certainty: Knowing what we legislate

Clarity over key legal terms is crucial to define the scope of a law and thus, to ensure legality and legal certainty. The principle of legal certainty is a general principle of EU law and part of the rule of law framework as defined by the Council of Europe. As pointed out by different scholars such as Lon Fuller, law simply cannot exist without legal certainty.[84] The European Court of Justice affirmed the element of clarity as essential for any legislation: "The principle of legal certainty requires that rules of law be clear and precise and predictable in their effect, so that interested parties can ascertain their position in situations and legal relationships governed by EU law".[85] Moreover, in the context of criminal law, legal certainty and clarity is necessary in order to establish a person's culpability (*mens rea*). If it is unclear whether a certain action even constitutes a crime, individual criminal responsibility cannot be established.

As highlighted earlier, there is no universal definition of disinformation that countries can rely on when designing such legislation. In case of the platform regulatory model, the respective

[83] András Sajó and Renata Uitz, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (Oxford University Press 2017) 415
[84] Lon Fuller, *The Morality that makes law possible* (Yale University Press 1969) 33-94. In: Lon Fuller, *The Morality of Law* (Yale University Press 1969)
[85] Case C-48/14 *European Parliament v Council of the European Union* [2015] ECR I-0048 para 45

laws do not use the term disinformation but refer to existing provisions in the Criminal Code to define "illegal conduct". Reference is not made to content per se but prohibited conduct. This means that the criminal law provisions allow for the prohibition of disinformation only in specific cases under specific circumstances. In other words, certain criteria must be met to be subsumed as a crime, which leads to legal unclarity. For example, in the case of the German *NetzDG*, it is unclear whether the provision on incitement to hatred only requires the provider to check whether hatred is incited or whether further criteria such as intention have to be established.[86] Moreover, many of the provisions are themselves considered overly broad and vague under international freedom of expression standards.[87] Additionally, it is unclear how to distinguish between "illegal content" and "manifestly" illegal content, the latter requiring the online provider to take down content within 24 hours. This violates international human rights standards. The UNSR has found any obligations that regulate or delete content on the basis of "vague and ambiguous criteria" incompatible with Art 19 of the ICCPR and the international right to freedom of expression.[88]

The state regulatory model does not refer to existing criminal law provisions but introduces a new offense. Here, the ambiguous term "false information"[89] or the claim or spread of "falsehood" or "a distorted truth"[90] is used as the basis for disinformation laws. These terms focus primarily on the distinction between truth and falsehood. However, is it really that easy to distinguish between what is true and false? Which criteria are used to determine falsehood? As Donald Trump's presidency showed in the United States, "fake news" is a highly relative term. Consequently, criticism was raised that disinformation laws such as the French Law on

[86] Sandra Schmitz and Christian M. Berndt, 'The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword?' (A Blunt Sword) [2018] SSRN, 24
[87] Article 19, 'France: Analysis of draft hate speech bill' (3 July 2019) https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/ accessed 13 June 2021
[88] Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, OL DEU 1/2017 [1 June 2017] 4
[89] French Law n° 2018-1202 against the Manipulation of Information
[90] Section 337 of the Hungarian Bill T/9790

the Manipulation of Information would give the state a monopoly over the truth.[91]

It seems that the inherent problem with this kind of legislation is the desire to forcibly legislate an area that, by definition, escapes certainty. As a consequence, such legislation and simplification creates false certainty and opens doorways to an arbitrary and abusive interpretation of the law.[92] It also leads to a chilling effect on freedom of expression as individuals may avoid controversial topics due to the uncertainty of what is permitted and what is not.[93]

2.3. Separation of powers: Who holds the truth?

The underlying vagueness of disinformation laws gives the respective decision-making body significant power over what is true or false. The platform regulatory laws transfer this assessment to the platforms who then decide whether a content is "illegal" or "manifestly illegal". This model raised serious concerns as states should avoid to delegate responsibility to companies as adjudicators of sensitive content. Effectively, this empowers corporate judgement. The UNSR stated in this regard: "States should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy".[94]

Such independence and impartiality is, by its very nature, not ensured by a private entity that is inherently business-driven and acts on the basis of its private interest and that, moreover, faces procedural obstacles such as an extremely short review window of 24 hours against more than

[91] Angelique Chrisafis, 'French MPs criticise "hasty and ineffective" fake news law' *The Guardian* (London, 8 June 2018) https://www.theguardian.com/world/2018/jun/07/france-macron-fake-news-law-criticised-parliament accessed 13 June 2021

[92] EDRi, Contribution to the examination of France's draft law aimed at combating hate content on the internet' (Brussels, 18 November 2019) 4 https://edri.org/wp-content/uploads/2019/11/20191118_EDRiCommentsEC_FrenchAvialaw.pdf accessed 13 June 2021

[93] Centre for Law and Democracy, 'Restriction on freedom of expression' (Briefing Note 2 of 12, IMS 2015) http://www.law-democracy.org/live/wp-content/uploads/2015/02/foe-briefingnotes-2.pdf accessed 18 March 2021

[94] United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2018] A/HRC/38/35 para 66

a billion posts per day.[95] In the context of the French Avia Law, which shares the key feature of a 24-hour notice-and-take-down system, the CC found the law to be unconstitutional. One major concern highlighted by the CC was the obligation to take-down manifestly illegal content within 24 hours as being too short of a period for the necessary evaluation of the content and its "manifest illegality", particularly as the timeframe also prevented any judiciary involvement in determining the legality of content.[96] Surprisingly, the CC also emphasised that manifestly illegal content does not have to be authorized by a judge per se.[97] What seems, however, crucial is the *independence* of the decision-making body and that any legislation restricting free speech is applied by an institution which is "independent of any political, commercial, or unwarranted influences in a manner that is neither arbitrary nor discriminatory."[98]

Aside from the question whether a non-judicial body can ensure the same level of independence and impartiality as the judiciary, legal training and in-depth knowledge of the respective national legislation and domestic case law is certainly needed in order to adequately assess the illegality of the content. Undoubtedly, interpreting the law and deciding on the legality of content is the primary role of the judiciary and a privatization raises constitutional concerns, in particular when touching upon fundamental rights such as freedom of expression. That being said, it would be an illusion to think that the judiciary is the solution to the problem of disinformation laws. A legislation that lacks legal certainty cannot be remedied by the decision of a judge, regardless of their legal expertise.

Among current disinformation laws, the French Law against the Manipulation of Information is the only legislation which foresees a fast-track authorization of content-removal by a judge. As part of an emergency procedure, an interim relief judge, which is the fastest possibility to

[95] See e.g. Facebook https://about.fb.com/company-info/ accessed 13 June 2021
[96] Conseil Constitutionnel, Décision n° 2020-801 DC of 18 June 2020 para 14ff
[97] ibid, para 19
[98] United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2011] A/HRC/17/27 para 24

grant justice, must decide within 48 hours on the truthfulness of the information and the intent of the speaker to manipulate public opinion. As part of the preliminary legislative review process, the CC issued a reservation of interpretation regarding the judge's assessment, namely that falsity must be demonstrated objectively and that the dissemination of such allegations or imputations must be artificial or automated, occur on a massive scale and be deliberate.[99] Additionally, the risk of influencing the election must be manifest and the judge assigned to the case must only apply the least infringing measures and meet the criteria of necessary, appropriate and proportionate measures.[100] This clarification did not, however, alter the initial set-up, which risks not only judicial overload, but the rapid deletion requirement makes the possibility to assess the context of the case almost impossible.[101] Due to the vagueness and ambiguity of the law, it leaves judges with the pressure, but also the discretionary power, to decide on the circulation of information during a highly sensitive election period. Such a construction increases the interest of a government to assert political pressure on the judiciary and thus makes the judiciary more vulnerable.

Similarly, the Hungarian provision leaves the decision on what constitutes "falsehood" or "a distorted truth" in relation to the state of emergency during the COVID-19 pandemic to the judiciary as well. In contrast to the French law, no timeframe of 48 hours is foreseen. Instead, a judge may sentence individuals to up to five years of prison if the false news were deemed to be "suitable for obstructing or preventing successful protection" of the public.[102] Given the lack of judicial independence in Hungary, a constitutional concern raised by numerous international bodies[103], this provision risks to be used against anyone who challenges the government's

[99] Conseil Constitutionnel, Décision No. 2018-773 DC du 20 décembre 2018 para 21
[100] ibid, para 23ff
[101] Judit Bayer and others, 'Disinformation and propaganda' (European Parliament 2019) 101
[102] Section 337 of the Hungarian Bill T/9790
[103] See for example the 2020 Rule of Law Report on Hungary by the European Commission, SWD(2020)316 final

actions during the pandemic.[104] Particularly, media outlets and journalists have voiced concerns as they fear that this law can be easily used against them. It is important to point out that this law has been passed in a climate where independent media is already largely curtailed.[105]

Concluding, these examples showcase the necessary caution to avoid sliding down a slippery slope and into the arms of a ministry of truth by criminalizing vague content-based concepts such as disinformation. Although the judiciary's involvement is an important constitutional safeguard to ensure the separation of powers, the danger from disinformation laws emanates less from the lack of judicial authority than the introduction of disinformation laws in the first place. Such wide discretionary power should not be transferred to any body, including the judiciary.

2.4. Criminalization of disinformation: Incentivizing over-blocking and censorship?

High penalties have a clear deterrent purpose. Therefore, the UNSR recommended in cases of free speech that states should "refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression".[106] This is in line with the case-law of the ECtHR which always looks at the nature and severity of the penalty as part of the proportionality test.[107] Regarding free speech cases, the ECtHR held that imprisonment will only be compatible with Art 10 ECHR in

[104] Kim Lane Scheppele, 'Hungary Is on the Edge of Dictatorship' (*Public Seminar*, 11 April 2020) https://publicseminar.org/essays/hungary-is-on-the-edge-of-dictatorship/ accessed 13 June 2021

[105] According to the World Press Freedom Index of Reporters Without Borders, Hungary ranks 89th out of 180 countries "with total indefinite control over the media under the coronavirus law" and facing "a degree of media control unprecedented in an EU Union member state". Reporters Without Borders, 'Countries – Index in the time of coronavirus' https://rsf.org/en/countries accessed 18 March 2021
Reporters Without Borders, 'Level of media control in Hungary is "unprecedented in an EU member state"' (5 December 2019) https://rsf.org/en/news/level-media-control-hungary-unprecedented-eu-member-state accessed 13 June 2021

[106] United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2018] A/HRC/38/35, 19

[107] *Ceylan v. Turkey* App no. 23556 (ECtHR, 8 July 1999) para 37

exceptional circumstances, "notably where other fundamental rights have been seriously impaired, as, for example, in cases of hate speech or incitement to violence".[108] In the case of France, a violation of the Law against the Manipulation of Information leads to one year imprisonment and a fine of €75.000.[109] In Hungary, the penalty is up to five years of imprisonment for spreading "false news".[110] Crucially, these penalties can target any individual, including journalists. They are not linked to further criteria such as inciting instability of the whole of society. This is in tension with the ECtHR's case-law which grants the press a particular high level of protection due to the journalist's fundamental role in a democratic society as a public watchdog. Even criminal penalties of relatively small fines such as €150 were considered potential censorship by the Court: "In the context of the political debate such a sentence would be likely to deter journalists from contributing to public discussion of issues affecting the life of the community."[111] In the case of disinformation laws, the risk of a chilling effect due to criminal penalties is heightened when facing legal uncertainty as highlighted earlier and when discussing controversial issues that might be contrary to the government's position.

In the case of the platform regulatory model, it is the high fees that social networks face for failing to block or remove content which have a potential chilling effect on free speech. This is due to economic reasons as it is simply cheaper for social networks to take down content in case of doubt. The German *NetzDG* imposes sanctions of up to € 500 million for companies[112], the French Avia Law up to € 20 million or 4% of the annual turnover for serious and recurrent failures to remove content[113] while the Austrian version imposes penalties of up to € 10 million

[108] *Mahmudov and Agazade v. Azerbaijan* App no. 35877/04 (ECtHR, 18 December 2008) para 50
[109] Art 1 *loi* n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information in conjunction with Art 112 of the Electoral Code (Code electoral).
[110] Section 337 of the Hungarian Bill T/9790
[111] *Lingens v. Austria* App no. 9815/82 (ECtHR, 8 July 1986)
[112] § 4 (2) and (4) NetzDG in conjunction with § 20 (2) 3 OWiG (Code on Administrative Offences)
[113] Marc Schuler and Benjamin Znaty, 'New law to fight online hate speech to reshape notice, take down and

for companies for recurrent failure of the moderation procedure.[114] Despite the lower penalty in Austria, this law also targets smaller online service communication providers with more than 100.000 users per year or an annual turn-over of € 500.000. Additionally, the national contact person may face penalties of € 10.000 in Austria and € 5.000 in Germany for failure of providing information.[115] A very practical question is who is willing to serve as a representative when faced with such fines.

Within an economic logic of costs and benefits, it is a structural mistake to sanction the failure of content deletion but not to provide protection against over-blocking.[116] Indirectly, the issue of criminalization also concerns the offenses that are listed as "illegal content" under the platform regulatory laws. Regarding defamation, for example, advocates of media freedom including the UNSR have urged states to repeal criminal defamation laws "in favour of civil laws as the latter are able to provide sufficient protection for reputations".[117]

2.5. The need for independent review and the right to appeal

A key constitutional safeguard is the right to appeal to ensure access to justice.[118] Any court whose decisions cannot be appealed would run the risk of acting arbitrarily.[119] This logic applies even more so if the decision-making body is not the judiciary in the first place.[120] In the case of

liability rules in France' (*Taylor Wessing*, 21 May 2020) https://www.taylorwessing.com/en/insights-and-events/insights/2020/05/new-law-to-fight-online-hate-speech-in-france accessed 13 June 2021

[114] § 10 (1) and (2) KoPl-G

[115] § 10 (4-6) KoPl-G; § 4 (2) and (4) NetzDG in conjunction with § 20 (2) 3 OWiG (Code on Administrative Offences)

[116] Mathias Hong, 'Das NetzDG und die Vermutung für die Freiheit der Rede' (*Verfassungsblog*, 9 January 2018) https://verfassungsblog.de/das-netzdg-und-die-vermutung-fuer-die-freiheit-der-rede/ accessed 13 June 2021

[117] United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2000] E/CN. 4/2000/63, para 52

[118] The right to appeal is expressly guaranteed by Article 2 Protocol 7 ECHR and Article 14.5 ICCPR in the criminal field, and by Article 8.2.h ACHR in general.

[119] European Commission for Democracy through Law (Venice Commission), CDL-AD(2016)007 [2016] para 105

[120] In Romania, where a public authority makes the decision on the illegality of content, an appeal or redress mechanism is not foreseen. This contradicts the rule of law requirement that any executive discretion must be controlled by judicial or other independent review. European Commission for Democracy through Law (Venice Commission), CDL-AD(2016)007 [2016] para 65-66

France and Hungary, where the decision is taken by a judge, this issue does not arise as the regular appeal procedure within the respective judicial system applies. However, the lack of a clearly outlined complaint and redress mechanism for individuals whose content has been taken down has been highly criticized in relation to the German *NetzDG* as well as the French Avia Law. The *NetzDG* requires the service provider to notify any person who submits a complaint about its decision including the reasons, but there is no adversarial process that would allow the person whose post is being complained about to challenge the potential blockage or removal of content.[121] Furthermore, the *NetzDG* does not include an obligation to publish the individual decisions on complaints and in case of failure to delete content, there is no further process for individuals foreseen in the *NetzDG*.[122] The explanatory notes to the *NetzDG* for setting fines made clear that the Act's aim is to only punish "systemic failure" of decision-making practice of social networks to avoid the risk of "over-blocking". This is not specified in the Act itself but the assessment of whether a systemic failure exists is overseen by a court.[123] The lack of appeal against the withdrawal request was also highlighted as problematic by the CC when declaring the Avia law as unconstitutional.[124]

In this regard, the Austrian model constitutes an improvement as it includes a process which allows the user whose content was blocked to appeal the provider's decision. Any person who issues a complaint must receive information about the procedure of the complaint mechanism, the decision and its basis and their right to appeal.[125] The same applies to the person whose post is being complained about. In case of an appeal, the platform is obliged to provide a review within two weeks.[126] If the outcome is unsatisfactory, a second-level redress can be brought in

[121] Of course, the user has the option to challenge the removal or blockage of content on social media platforms such as Facebook in front of a civil court but a redress procedure as such is not foreseen under the NetzDG.
[122] There is, however, a judicial procedure foreseen for providers who face administrative penalties, see § 4 (5) NetzDG.
[123] Sandra Schmitz and Christian M. Berndt, 'A Blunt Sword?' [2018] SSRN, 25
[124] Conseil Constitutionnel, Décision n° 2020-801 DC of 18 June 2020 para 7
[125] § 3 (3) 2 KoPl-G
[126] § 3 (4) KoPl-G

front of the arbitration body of the telecom and media regulator RTR.[127] However, a court is not involved in the appeal procedure to assess whether the content has actually been illegal and fundamental rights were duly considered.

Furthermore, significant improvements were made in terms of transparency obligations as the reporting mechanism, which also exist in Germany and France, requires platforms to disclose more detailed information regarding the content moderation process and statistics (e.g. the training of content moderators and technical equipment).[128] In contrast to Germany, platforms operating in Austria must report on any content blockage or deletion even if based on internal regulation procedures. This allows for a much more comprehensive understanding of how, when and on which basis content is taken down by private platforms, thus also increasing accountability.

Concluding, judicial review and the right to appeal is a key feature to prevent abuse of power and to ensure the accountability of the decision-making body. Any form of government pressure on the decision-making body to censor escapes scrutiny without an independent review mechanism in place. Apart from a non-judicial redress mechanism in Austria, this is not ensured in Germany and in the French Avia draft law. However, general transparency measures as highlighted in Austria increase the overall accountability of private decision-making bodies.

Conclusion: Abuse of power and implications for liberal democracy

The role and impact of online service providers, namely social media platforms, on public discourse and democratic processes has changed dramatically over the last century.[129] Without question, platforms should be held responsible for the actions they perform. In recognition of

[127] § 8 (2) KoPl-G
[128] § 4 KoPl-G
[129] See in recognition of the role of social networks for the implementation of constitutional principles such as the pluralism of political parties a decision of the Court of Rome against censorship: *CasaPound v. Facebook* App no 80961/19 (Court of Rome, 29 April 2020)

the need to better regulate the digital single market, the European Union is currently discussing updates of the e-Commerce Directive by introducing the Digital Services Act.[130] In light of that, this thesis is less concerned with the clear need to tackle competition and monopolization issues than to warn of a continuous spread of controlling content by introducing disinformation laws in the name of democracy.

The aim of the thesis was to provide an attempt at classifying how EU countries regulate this area. This should allow for a better overview of the developments and different models of disinformation laws in Europe. Secondly, I aimed to highlight the constitutional risks of such disinformation laws, not just for censoring certain kinds of speech but their potential for abuse of power – may it be public or private. I argued that the valid concerns about disinformation campaigns should not blind the EU and its Member States to see how such regulations may open Pandora's box. Borrowing from legal expert Jacob Mchangama:

> "The fact that the truthfulness of news, reporting and information has always been contested does not necessarily mean that the reaction to current developments in the digital age should be shrugged off as a moral panic without any merit. But it should caution decision makers tempted to adopt draconian measures without fully understanding the likely consequences. The preceding history reveals the dangers of putting governments and institutions in charge of defining truth and error."[131]

As this thesis shows, the loopholes for abuse start with the lack of legal certainty regarding the concept of disinformation, which gives the respective decision-making body great discretion when determining the illegality of content, often without judicial supervision. The criminalization and high sanctions heighten the chilling effect on free speech. While Austria

[130] European Commission, 'The Digital Services Act package' https://ec.europa.eu/digital-single-market/en/digital-services-act-package accessed 13 June 2021

[131] Jacob Mchangama, 'Fake News is Old News' (Quillette, 25 August 2017) https://quillette.com/2017/08/25/fake-news-old-news/ accessed 13 June 2021

learned some lessons based on the criticism of the German *NetzDG* and the decision of the CC regarding the Avia Law (e.g. imposing lower sanctions, the obligation for platforms to ensure an internal redress mechanism of platforms and higher transparency standards), the general set-up did not eliminate the most important concern around such laws, that is the decision-making on the (il)legality of criminal offenses by platforms and the very short timeframe of 24 hours.

Regarding the state regulatory model, the use of the vague term "false information" puts the judge in the French model in a similar position as the platforms, which – despite the needed judicial independence and legal expertise – still faces the problem of legal uncertainty. Moreover, the case of Hungary shows how fast governments act on exerting control over the truth in times of emergencies. Once we allow the state or platforms to censor public debate by criminalizing disinformation, we enter a slippery-slope, and it will become very hard to argue when the empowered institution crossed a line. The bitter experience of emerging illiberal democracies at the heart of Europe should remind the EU of the importance of procedural and substantive safeguards that do not depend on the trust in specific governments or legal traditions. Thankfully, the proposed Digital Services Act by the European Commission does not follow either of the presented disinformation model. Given the constitutional risks described in this thesis, the EU should carefully consider whom it wants to put in charge to define the truth in society.

Bibliography

Books, Journal articles and Reports:

András Sajó and Renata Uitz, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (Oxford University Press 2017)

Giovanni Pitruzzella and Oreste Pollicino, *Disinformation and Hate Speech: A European Constitutional Perspective* (Bocconi University Press 2020)

High level Group on fake news and online disinformation, A multi-dimensional approach to disinformation (European Commission 2018)

Judit Bayer and others, 'Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States' (European Parliament 2019)

Lon Fuller, *The Morality of Law* (Yale University Press 1969)

Olga Robinson Alistair Coleman and Shayan Sardarizadeh, 'A Report of Anti-Disinformation Initiatives' (Oxford Technology & Elections Commission 2019) https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/08/A-Report-of-Anti-Disinformation-Initiatives

Sandra Schmitz and Christian M. Berndt, 'The German Act on Improving Law Enforcement on Social Networks (NetzDG): A Blunt Sword?' [2018] SSRN

Legal texts:

Council Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) [2000] L 178/1

Hungarian Bill T/9790

Kommunikationsplattformen-Gesetz (Austrian Communications Platform Law)

loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (French Law n° 2018-1202 against the Manipulation of Information)

Netzdurchsetzungs-Gesetz (German Network Enforcement Act)

Romanian Presidential Decree No. 195/2020

Proposition de loi, adoptée par l'Assemblée nationale dans les conditions prévues à l'article 45 alinéa 4 de la Constitution visant à lutter contre les contenus haineux sur internet le 13 mai 2020, T.A. n° 419 (Avia Law n° 419)

Case law:

CasaPound v. Facebook App no 80961/19 (Court of Rome, 29 April 2020)

Case C-48/14 *European Parliament v Council of the European Union* [2015] ECR I-0048

Ceylan v. Turkey App no. 23556 (ECtHR, 8 July 1999)

Chaplinsky v. New Hampshire 315 U.S. 568 (1942)

Conseil Constitutionnel, Décision n° 2020-801 DC of 18 June 2020

Delfi AS v Estonia, App no 64569/09 (ECtHR, 16 June 2015)

Handyside v United Kingdom App no 5493/72 (ECtHR, 7 December 1976)

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

Mahmudov and Agazade v. Azerbaijan App no. 35877/04 (ECtHR, 18 December 2008)

Schenk v. United States, 249 U.S. 47 (1919)

United States v. Alvarez 567 U.S. 709 (2012)

Whitney v. California 274 U.S. 357 (1927)

Other documents:

2020 Rule of Law Report on Hungary by the European Commission, SWD(2020)316 final

European Commission for Democracy through Law (Venice Commission), CDL-AD(2016)007 [2016]

'Leitlinien zur Festsetzung von Geldbußen im Bereich des Netzwerkdurchsetzungsgesetzes (NetzDG)' (*Federal Ministry of Justice and Consumer Protection,* 22 May 2018) https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/NetzDG_Bu%C3%9Fgeldleitlinien.html

Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, OL DEU 1/2017 [1 June 2017]

Republic of Austria, Parliamentary correspondence Nr. 1391 of 10 December 2020 https://www.parlament.gv.at/PAKT/PR/JAHR_2020/PK1391/index.shtml

United Nations Human Rights Committee, General Comment No. 34 [2011] CCPR/C/GC/34

United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2000] E/CN. 4/2000/63

United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2011] A/HRC/17/27

United Nations, Report of the Special Rapporteur on the Promotion and Protection of the Right on Freedom of Opinion and Expression [2018] A/HRC/38/35

Website of the Minister of Internal Affairs https://www.mai.gov.ro/precizare-23/

Online Articles:

Angelique Chrisafis, 'French MPs criticise "hasty and ineffective" fake news law' *The Guardian* (London, 8 June 2018) https://www.theguardian.com/world/2018/jun/07/france-macron-fake-news-law-criticised-parliament

Article 19, 'France: Analysis of draft hate speech bill' (3 July 2019) https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/

BBC News, 'Coronavirus: Hungary votes to end Viktor Orban emergency powers' (16 June 2020) https://www.bbc.com/news/world-europe-53062177

Centre for Law and Democracy, 'Restriction on freedom of expression' (Briefing Note 2 of 12, IMS 2015) http://www.law-democracy.org/live/wp-content/uploads/2015/02/foe-briefingnotes-2.pdf

Daniel Funke and Daniela Flamini, 'A guide to anti-misinformation actions around the world' (*Poynter*) https://www.poynter.org/ifcn/anti-misinformation-actions/

EDRi, 'Contribution to the examination of France's draft law aimed at combating hate content on the internet' (Brussels, 18 November 2019) https://edri.org/wp-content/uploads/2019/11/20191118_EDRiCommentsEC_FrenchAvialaw.pdf

Emma Charlton, 'How Finland is fighting fake news – in the classrom' (*World Economic Forum*, 21 May 2019) https://www.weforum.org/agenda/2019/05/how-finland-is-fighting-fake-news-in-the-classroom/

European Commission, 'The Digital Services Act package' https://ec.europa.eu/digital-single-market/en/digital-services-act-package

epicenter.works, 'First Analysis of the Austrian Anti-Hate Speech Law (NetDG/KoPlG)' (*EDRi*, 10 September 2020) https://edri.org/our-work/first-analysis-of-the-austrian-anti-hate-speech-law-netdg-koplg/

Georg Mascolo and Ronen Steinke, 'Bedenken in Bellevue' *Süddeutsche Zeitung* (Munich, 17 September 2020) https://www.sueddeutsche.de/politik/hate-speech-hasskriminalitaet-gesetz-steinmeier-1.5034929

iate, https://iate.europa.eu/entry/result/3576921

Jacob Mchangama, 'Fake News is Old News' (Quillette, 25 August 2017) https://quillette.com/2017/08/25/fake-news-old-news/

Kim Lane Scheppele, 'Hungary Is on the Edge of Dictatorship' (*Public Seminar*, 11 April 2020) https://publicseminar.org/essays/hungary-is-on-the-edge-of-dictatorship/

Lida Filippakis, 'Croatian anti-hate law might lead to censorship, some experts fear' *Independent Balkan News Agency* (Bourgas, 19 January 2018) https://balkaneu.com/croatian-anti-hate-law-might-lead-to-censorship-some-experts-fear/

Mathias Hong, **'**Das NetzDG und die Vermutung für die Freiheit der Rede' (*Verfassungsblog*, 9 January 2018) https://verfassungsblog.de/das-netzdg-und-die-vermutung-fuer-die-freiheit-der-rede/

Marcel Gascón Barberá, 'Romania's Drive to Censor 'Fake News' Worries Activists' (*Balkan Insight*, 27 April 2020) https://balkaninsight.com/2020/04/27/romanias-drive-to-censor-fake-news-worries-activists/

Marc Schuler and Benjamin Znaty, 'New law to fight online hate speech to reshape notice, take down and liability rules in France' (*Taylor Wessing*, 21 May 2020) https://www.taylorwessing.com/en/insights-and-events/insights/2020/05/new-law-to-fight-online-hate-speech-in-france

Media Legal Defence Initiative, Module 3: Criminalisation of online speech (2020) https://www.mediadefence.org/ereader/wp-content/uploads/sites/2/2020/06/Module-3-Criminalisation-of-online-speech.pdf

OSCE, 'Coronavirus response bill should not curb freedom of information in Romania, stresses OSCE Media Freedom Representative' (30 March 2020) https://www.osce.org/representative-on-freedom-of-media/449380

Peter Walker, 'New national security unit set up to tackle fake news in UK' *The Guardian* (London, 23 January 2018) https://www.theguardian.com/politics/2018/jan/23/new-national-security-unit-will-tackle-spread-of-fake-news-in-uk

Reporters Without Borders, 'Countries – Index in the time of coronavirus' https://rsf.org/en/countries accessed 18 March 2021

Reporters Without Borders, 'Level of media control in Hungary is "unprecedented in an EU member state"' (5 December 2019) https://rsf.org/en/news/level-media-control-hungary-unprecedented-eu-member-state

'Repressive laws, prosecutions, attacks… Europe fails to shield its journalists against the abuse of the COVID-19 crisis' (*Reporters Without Borders*, 8 April 2020) https://rsf.org/en/news/repressive-laws-prosecutions-attacks-europe-fails-shield-its-journalists-against-abuse-covid-19

'Spain and Russia agree to set up joint cybersecurity group' *AP News* (New York, 6 November 2018) https://apnews.com/article/00061a7eb8814fc38ebef55880c3fd18

'Sweden to create new authority tasked with countering disinformation' *The Local* (Stockholm, 15 January 2018) https://www.thelocal.se/20180115/sweden-to-create-new-authority-tasked-with-countering-disinformation