HUMBOLDT-UNIVERSITÄT ZU BERLIN


**Forschungsinstitut für Recht und digitale Transformation**

Working Paper Series

2021

# Accountability of Social Media – A European Perspective

Prof. Dr. Martin Eifert, LL.M. (Berkeley)

Working Paper No. 10

Zitiervorschlag: Martin Eifert, „Accountability of Social Media – A European Perspective", Working Paper No. 10 des Forschungsinstituts für Recht und digitale Transformation (2021).

# Accountability of Social Media – A European Perspective*

– Martin Eifert* –



* Lecture, given at the opening of the Legal Ground Institute, Sao Paulo, 4 May 2021.

* Martin Eifert holds a chair of Public Law at the Faculty of Law, Humboldt-Universität zu Berlin.

# I. Introduction

Thank you very much for the kind introduction. It is an honor and a great pleasure for me to give today's lecture on the responsibility of online intermediaries.

I will focus on social media. This is only one type of intermediaries but a very important one. Social Media has become a crucial means for individual self-realization as well as for community building and public deliberation. Its regulation has got into the focus of legislators all over the world and is intensely discussed among politicians and academics.

Until recently, Social Media was typically covered by the broad liability shields for all internet intermediaries which had been enacted decades ago in the EU, the US and elsewhere. The liability shields reflected the predominant notion to avoid anything that could stifle innovation at the dawn of promising new internet services.

In the meantime, however, social media has gained tremendous significance in social life, and its economic particularities has left only very few and powerful providers. Social media's impact and its market structure have fundamentally changed the way we think about its responsibilities.

Two developments are driving the discussion. Firstly, social media not only fosters public discourse by side-stepping established mass-media gatekeepers, it also enables users to disseminate hate-speech, misinformation and all kinds of illegal content. The reach and dynamics of platforms amplify the harm resulting from it. As a consequence, the focus has shifted from liability exemptions to designing an adequate scheme of due diligence obligations. The major challenge for any approach to this problem is to prevent harm on the one side without having a chilling effect on free speech or stimulating overblocking by platform operators on the other side. This regulatory challenge obviously implies difficult human rights questions.

Secondly, social media do not only sidestep traditional mass media gatekeepers, they have become gatekeepers themselves. Their terms and conditions and algorithms define what can be said and heard, and their power to remove content and to block users has significant influence on public discourse. Many of us are concerned because we do not want private companies to exclude important controversial issues from public deliberation. Again, difficult human rights issues are raised. To what extent are the platforms free to set standards as they like as part of their business model and to what extent are they required to take the right to freedom of expression of their users into consideration?

Before starting our exploration of the regulatory issues, we should make sure that we have a shared notion of the situation we are talking about. Therefore, we will first of all have a closer look at Social Media Communication and the problems entailed with it.

Afterwards, we will roughly summarize the path legislators in Europe have taken in order to address at least some of these problems before we turn to the different issues and the respective regulatory approaches. We will focus on three topics, starting with illegal content and the responsibilities related to it. Then, we will examine questions of private content moderation beyond such illegal content. Eventually, we will ask how to adequately deal with systemic risks for liberal democracies that may result from effects like echo-chambers or content which is not illegal, but nevertheless harmful. We will see that a common thread running through all legislation is the approach to set a framework for self-regulation, leaving the operators the more leverage, the more the issues are free-speech-sensitive. Procedural and organizational requirements are the core of social Media regulation.

## II. Social Media: What is the Problem – and how to deal with it?

### 1. The Triangular structure of platform communications

Let us start with a brief description of the phenomena and problems we are facing.

Social Media have blurred the line between mass communication and individual communication. They have created a network of users that are generating and consuming content on social media platforms which allow for easy access and high connectivity. For analytical purposes, it is helpful to focus on the triangular structure which is underlying platform communications.

We have a great number of users, all of whom can produce and consume content - but in each single act of communications every user is either sender or recipient of content. The Social Media platform is providing the infrastructure for all communications. It enables users to communicate online by providing easy access, making available tools for producing content, and generating opportunities to connect with other users. As Social Media foster content-generation and connectivity, they also amplifying all effects of user content.

The easily accessible infrastructure has created a space for communication that contains an unprecedented amount of speech. Additionally, persistence, reach and dynamics of online-speech allow for far-reaching effects. Online content is available without timelimit, can reach an unlimited number of recipients and takes part in the dynamics of online communication - we all know phenomena like shit storms or content going viral.

If this is the setting – what is the problem?

The first problem is that the unprecedented amount of speech also contains an unprecedented amount of harmful speech and that social media also amplify the harm entailed with it. The most obvious harm is infringement of private rights or violation of laws which also protect public interests, in other words: illegal content.

The second problem is closely linked to the market structure. Control and curation exercised by social media operators become a problem if there are only very few of them. The more public deliberation takes place on these very few social media platforms, the more influential are decisions of social Media operators about content and information flows on the platform. Economic market power turns into impact on public discourse.

Thirdly, we are concerned about more indirect, systemic effects. Speech which is not illegal, such as disparaging messages which do not qualify as defamation or Fake-News, can nevertheless be used to intimidate vulnerable groups, to erode trust in government institutions, to disparage groups, opinions, office holders, or science. Even though the single act of communications is not illegal and should not be illegal as it does not result in immediate harm, the sum and dynamics of such communications can nevertheless result in serious harm. This is especially the case if certain patterns apply. The amount of such speech can be enhanced by fake-accounts or bots and surpass any capacity of public discourse to make effective corrections; The emergence of communities which echo such communications over and over again can foster radicalization; and the systematic targeting of individuals or vulnerable groups with such speech can silence their voices. In short: Such speech, all the more if combined with illegal speech, may undermine preconditions of public deliberation and democratic governance. This is a systemic effect.

### 2. Crucial regulatory issues: redress against primary infringer, liability of intermediaries for illegal content, private content moderation, and systemic effects

The regulatory issues, of course, reflect the problems I have just mentioned.

It is first of all private users, not the platforms, who infringe private rights. But it is difficult and often impossible for aggrieved parties to gain redress from the primary infringer, especially because his or her name and address is in most cases not known by the aggrieved party. This

spotlights the responsibility of platforms for illegal user-content. They are a secondary infringer as they contribute to the harm by disseminating the illegal content. The regulatory question is, if or to what extent they should be required to use their control over content on the platform to prevent dissemination of illegal content. In other words, we are talking about content moderation by Social Media operators which is legally imposed.

The second problem relates to another prong of content moderation, the one which is based on so-called community standards and exercised by platforms irrespective of any legal framework. Community standards are embodied in general terms and conditions of the contracts between social media operators and their users. They define – inter alia – which content must not be uploaded. Platforms are enforcing the terms and conditions by removing content or even by blocking accounts in case of violations of the community standards.

It is important to notice that the scope of the community standards is overlapping with the legal order of the states but is not fully consistent with it. Some Social Media - such as facebook - include a clause which forbids to upload illegal content but the clause is accompanied by its own set of restrictions, other platforms do not have any reference to state law. Divergence between the legal order and community standards results mainly from two factors and their complex interplay: Firstly, a legal culture in Social Media that is still very much rooted in US notions of free speech and secondly the incentive structure which fosters a design that makes all users feel as comfortable as possible. Whatsoever the diverging area between legal content and community standards precisely looks like, this is the area in which private power, namely the power of social media operators, is constraining speech. The regulatory issue is whether or to what extent social media operators are free and should be free to set community standards as they wish.

The private power, however, is ambivalent. Some people call it private censorship as private companies prevent the dissemination of legal content. From this perspective, the community standards withdraw content from public deliberation and they prevent a robust public discourse without having sufficient legitimacy to do so. On the other hand, private content moderation is seen as an appropriate means to regulate harmful speech which – for reasons of constitutional protection of free speech - could not be regulated by the state. This is particularly the case with regard to the problem of systemic effects of some kinds of speech such as not prohibited variations of hate speech.

### 3. Fundamental rights implications

Let us conclude our first overview by briefly introducing the fundamental rights perspective on the problems and regulatory questions.

It goes without saying that any government regulation of social media must be aware of the fundamental rights of social Media operators. It is also obvious that it must be aware of negative effects on free speech. Regardless of how it is framed in constitutional doctrine, government regulation must not inhibit free speech unless the restrictions serve a legitimate goal and are tailored in a proportionate manner. German as well as European Constitutional law acknowledge the fundamental character of free speech for any liberal democracy and emphasize the sensitivity of free speech to detrimental effects. Therefore, traditional doctrines make it very difficult to outweigh free speech in a balancing test. Law of defamation and other forms of illegal content are recognized limitations of free speech but any enforcement by government or through legally designed mechanisms shall not have a chilling effect. If we apply the chilling effect doctrine to social media regulation, it certainly means that any approach should be eager to avoid incentives for overblocking of user content.

Potential infringements of private rights or committed criminal offenses like incitement to racial hatred by user-generated content are on the opposite side of the scale and are limiting the right to freedom of expression. In face of the amplification of harm, resulting from reach and dynamics of social Media interaction, we should not just apply traditional doctrinal rules but need to

reconsider the balance between freedom of expression and such private and public interests. When asking for the appropriate balance, we should also consider potentially existing constitutional obligations to protect such interests.

I hope that we have now a shared notion on the phenomena and the problems we are talking about. Let us turn to the answers which have been given thus far by European Legislators. I will start by giving you an overview of the developments so that you get an impression of the general approach and the momentum of European Legislation before describing in more detail the approaches to each of the problems which we have identified.

## III. Development of the legislative framework: From Liability shields to schemes of responsibility

### 1. Point of Reference: Limited Liability

From the outset of any specific regulation of intermediary services, limitations of liability have been a prominent feature. At the dawn of internet services, liability privileges were almost all regulation that was enacted. Legislators worried that liability could stifle innovation and were eager to facilitate the fast development of services such as social media platforms, a goal shared by the famous Section 230 of the Communications Decency Act in the US and, more important for us, Arts. 14 and 15 of the e-Commerce Directive in the EU. All liability exemptions, however, fell short of excluding all kinds of liability. While any general obligation to monitor user-generated content stored or transmitted is explicitly ruled out by Art. 15 E-Commerce-Directive, the liability exemption for user-generated content in Art. 14 e-commerce Directive is conditioned. More precisely, it is conditioned on expeditious removal of illegal content upon obtaining actual knowledge of it. From the perspective of aggrieved parties, this boils down to injunctive relief after notice but no compensation for any damage occurred.

I put so much emphasis on the liability exemptions because as a general rule, they have been upheld in any general European Regulation since then[†]. At first glance, this is surprising, because at least the nowadays dominant platforms definitely do not need an "innovation waiver" any more. Another justification for this approach is (still) valid, though. Social Media communications is done at scale. We are talking about billions of user-generated items per day on the big platforms such as facebook or twitter. Considering the sheer amount of speech, monitoring requires automatic means. Automatic means - at least by now – can not handle many of the context-sensitive evaluations that determine e.g. defamatory speech. If monitoring had to be made by human beings, however, it would become very costly if not impossible. Therefore, a due diligence obligation to monitor all content would render the business unprofitable or even unfeasible. It would be the end of the internet as we know it. In other words: The liability limitation lives on as the result of a proportionality test. The fundamental right of providers to conduct business is constrained by liability rules and a liability-regime which does not condition liability upon knowledge or awareness of a specific illegal content would not be proportionate as it would in fact shut down the operators. And after all, platforms only facilitate and amplify the harmful activities of their users, and do not commit the wrongs themselves.

### 2. Development in EU Member States

Notwithstanding the persisting justification for limited liability, a need for increased responsibility of Social Media has become evident. The E-Commerce-Directive, enacted as early as 2000, could not cope with the problems that were emerging as Social Media became more and more

[†] The only exception thus far is Art. 17 of the Copyright DSM Directive (2019/790) Directive on copyright and related rights in the Digitial Single Market) which is an explicit exception of Art. 14 E-Commerce Directive (Art. 17 (3) Copyright DSM Directive.

important and powerful. As a first step, national governments such as the German government as well as the European Commission started a dialogue with the powerful social Media operators in order to stimulate voluntary action against pressing issues. Especially Hate Speech had become a very pressing issue in times of intense discussion about migration politics. Even though there was some improvement on the side of at least some major Social Media, the overall results were not satisfying. Consequently, the German government - additionally fueled by an election campaign – introduced the Network Enforcement Act, which entered into force in 2017 after a fierce debate.

**a) German Network Enforcement Act**

The basic elements of the Network Enforcement Act have heavily influenced all European regulatory efforts ever since and are worth being described.

The core-element are procedural and organizational obligations for operators of Social Media platforms which comprise three requirements: Firstly, an easily accessible procedure to give notice about illegal content, secondly, a proper evaluation of incriminated content and thirdly, timely removal of illegal content within at least seven days, or in case of manifestly illegal content within 24 hours. The evaluation of the content can be done by the operator itself or it can be handed over to a recognized independent self-regulation institution.

Compliance with the statutory requirements is enforced by public administration and violations could lead to regulatory fines.

Apart from requiring a compliance-system, the Network Enforcement Act imposes rather detailed transparency obligations on Social Media operators. They have to report and publish biannually about the amount and types of unlawful content on their platform as well as about their measures in handling it.

Having roughly described the content, it is also important to underline what is deliberately not regulated by the Act. It refrains from establishing a liability system and instead builds on the existing liability of media platforms for interference as a transmitter of illegal content. The legal consequence of this liability is always removal of illegal content, not compensation for damages. It is also noteworthy that the Network Enforcement Act applies only to large networks (more than 2 Mio registered users in Germany) and addresses only some illegal content, namely already existing criminal offenses which are listed in the Act.

Not surprisingly, the discussion about the Act has continued in politics and academia. The paramount issue were (and still are) Free Speech concerns. Many commentators argue that the time limit, combined with public enforcement and fines in case of violations, is a strong incentive for operators to remove content if in doubt about its legality, so that the regulatory framework results in overblocking. This argument, however, does not adequately reflect the legal framework. It can be argued – and I personally share this point of view - that the requirements on complaint-handling comprise all aspects of proper evaluation and that under-blocking is as much subject to potential fines as it is overblocking. On this assumption, the incentive structure would be balanced with regard to any decision about removal. Furthermore, fines can´t be based on misjudgements in individual cases of content evaluation. Fining requires systematic failure in content evaluation and a respective court decision on the misjudgements, so that any individual judgement is unburdened. In 2020, we have conducted an evaluation of the Act and we could not find any evidence or indication for overblocking. Nevertheless, free speech is such a vulnerable value that it is easy to make a case for an improvement of safeguards.

Reflecting the ongoing discussion, amendments to the Network Enforcement Act have been introduced in parliament and will probably be enacted very soon. The most important amendments from the perspective of our topic are the following: Operators will be required to give reasons for their decisions upon removal of incriminated content and either party shall be given the opportunity to object to the decision and to require its re-evaluation. Additionally,

independent out-of-court dispute resolution panels will be set up and will decide on the reevaluation-decision on condition that both parties agree.

**b) Similar Approaches in France and Austria**

The German Network Enforcement Act is the most prominent regulation in Europe and the only one which is in force right now. France enacted a similar law in May 2020, called "loi Avia" but it was invalidated by the French Constitutional Court (Conseil Constitutional) just one month later. One major reason was the fact that according to this statute, operators had to decide within 24 hours with respect to some illegal content and could be fined on the basis of not removing a single item of content despite its illegality. It is quite obvious that this was a strong incentive for overblocking. Austria has drafted a law on Social Media Platforms ("Kommunikationsplattformgesetz") which is very much along the lines of the NetzDG and which has already passed the notification procedure with the EU.

**c) EU: Digital Services Act Proposal and other EU legislation**

Against the backdrop of the just described increasing activity among the member states, an obviously outdated European regulatory framework, and the ambitions of its digital single market strategy, the EU-Commission presented its proposal for a legislative package just 5 months ago.

It consists of the Digital Markets Act (DMA) and the Digital Services Act (DSA). The DMA focuses on economic power of internet-platforms, whereas the DSA focuses on the mitigation of externalities resulting from illegal content, together with the protection of the public interest in an open and free public discourse. By and large, the DSA resembles the approach taken by the German Network enforcement act but modifies it in important aspects.

At the outset, the DSA confirms the general limitations on the liability of internet service providers known from the e-Commerce Directive. The vast majority of its provisions, however, set up a regulatory framework with a special focus on so-called "online-platforms" which include Social Media platforms.

Such online platforms are subject to procedural and organizational requirements which first of all aim to ensure an effective notice-and-action mechanism for illegal content.[‡] The DSA does not only require a decision on the incriminated content in a timely, diligent and objective manner[§] including a statement of reasons, but also sets up a two-step complaint-handling system[**]. The latter starts with an internal review procedure[††] and provides for an obligatory subsequent out-of-court dispute settlement mechanism[‡‡]. Very large platforms with at least 45 million active recipients in the EU also need to engage in the assessment of systemic risks stemming from their services[§§] and undergo independent audits[***].

In addition to the elaborated complaint-handling procedures, the DSA establishes various transparency obligations for Social Media platforms, many of which relate to content regulation and risks for the public interest. Transparency obligations comprise the duty to include information on content restrictions imposed by the terms and conditions[†††] and also reporting

---

[‡] Art. 14, 15, 19, 20 DSA

[§] Art. 14 (6) DSA

[**] As for the complaint-handling system and a number of transparency obligations, very small platforms are exempted (Art. 16 DSA).

[††] Art. 17 DSA

[‡‡] Art. 18 DSA

[§§] Art. 25–28 DSA

[***] Art. 28 DSA

[†††] Art. 12 DSA

requirements about the means and measures of content moderation.[###] Finally, the DSA introduces public enforcement of the regulation including the power of public authorities to impose penalties such as fines.[§§§]

As a result of limited competences of the EU regarding the law of torts and criminal law, the DSA lacks genuinely European liability rules and any definition of what constitutes "illegal content" but instead simply refers to the law of the Member States[****]. Such strictly formal and uniform definition of illegal content forecloses the possibility to define especially harmful illegal content to be controlled and deleted as a priority by the platforms.

In sum, compared to the Network Enforcement Act, we can observe an even more elaborate system of decision-making on incriminated content, the emergence of some rules on terms and conditions, and the adoption of public enforcement. However, we do not have specified time limits for removal.

The brief survey has shown that Europe has entered a path towards increased responsibility of Social Media operators. The crucial element is a notice and action mechanism which is embedded in a system of review procedures. The legislative means are procedural and organizational requirements which set up and design such system in a way that protects free speech and avoids overblocking. We have also observed first and very cautious regulatory steps which address the terms and conditions of Social Media Operators and some measures which help to clear up potential systemic effects. Having mapped the path, it is time for an exploration of some crucial questions.

## IV. Illegal Content: Unfolding Notice and take-down Regimes

The attempt to hold platforms accountable for user-generated content turns out to be anything but trivial. Notice and take action is a convincing approach but if it is not more precisely designed by law, it provides a very strong incentive for operators to avoid getting any knowledge of illegal content. Not surprisingly, legislators have made significant efforts to make Notice-and Take-Down work.

### 1. Making Notice-and-take-down work

The German Network Enforcement Ac as well as the Austrian Law on Communications Platforms and the proposed European DSA require Social Media Platforms to provide mechanism which allow to notify them of specific harmful content on the platform. These mechanisms shall be easily recognizable, easy to access, and user friendly. In addition, the DSA-Proposal introduces so-called "trusted flagger"[††††] a status which can be awarded upon application to entities which have special expertise, represent collective interests and carry out their activity for the purpose of submitting notices. The notices of such trusted flaggers need to be processed with priority and without delay by the platforms. With regard to such notification-requirements the experience of the last years in Germany has shown, that close monitoring is necessary to ensure the implementation. The requirements of the German Network Enforcement Act will be further specified in its amendments and it remains to be seen whether the DSA-regulation and maybe some specifications in codes of conduct will be sufficient to achieve its goals.

The crucial element of notice and take down is of course the evaluation of the content which has been brought to the attention of the platforms. Regulations require an "effective and transparent

[###] Art. 13, 23 DSA

[§§§] Art. 42, 59 et seq. DSA.

[****] Art. 2 (g) DSA

[††††] Art. 19 DSA

procedure" (German Network Enforcement Act) or a decision in a "timely, diligent and objective manner" (EU DSA-Proposal). Given the uncountable varieties of content, it is clear that there can hardly be more precise regulations on the decision as such. Once again, procedural safeguards are the most promising approach to ensure proper decisions. Four aspects should be mentioned. First, in face of the amount of content, automatic means are unavoidable even though they can not be sufficient. The regulations ensure transparency about the fact that such means are used, but they do not address the algorithms as such. However, the more general discussion on accountable AI should also lead to more specific regulations. Second, evaluation of context depends heavily on legal and cultural expertise in the field of media. Social Media are operated by tech-firms which have such knowledge only to a limited extent. Regulation needs to ensure that they acquire adequate expertise. Current approaches are using indirect ways such as making the internal procedures subject to reporting (Network enforcement Act) and requiring a statement of reasons which is to be communicated also to the user who generated the content (NEA and DSA). An alternative way to ensure proper evaluation is setting up specialized entities which handle the notices for one or even many tech firms. This approach is also included in the Network Enforcement Act which makes it possible to transfer content-evaluation to an independent entity. The advantage of this alternative is independent review of content by experts, its disadvantage is the fact that Social Media operators do not regard it as part of their own business any more so that there is very little incentive to consider potential negative effects already in the development of the algorithms which shape the online environment.

Finally, in many cases, evaluation of the content is context-sensitive and the user who generated the content has presumably valuable information on this context. Is has been discussed to make compulsory some participation of the user who generated the content. On the other hand, as a standard procedure this would be cumbersome and time-consuming. In the end, we do not find such procedural requirements. Instead the user has to be informed about the decision and its reasons so that he or she can file a complaint in case of unjustified removal. In other words: A complaint-handling mechanism which allows for subsequent correction seems to become the standard as a safeguard against unjustified removal.

Let us turn to those mechanisms.

## 2. Redress-mechanisms for unjustified measures as free speech safeguards

According to the DSA-Proposal, online platforms must offer an effective internal complaint-handling system to the author of the incriminated content that must be easily accessible and user-friendly. The platform is bound to deal with the complaints of authors in a timely manner and it must not rely exclusively on algorithms. Where the platform finds that the complaint is well-founded it must reverse its decision and reinstate the incriminated content. In that case, the remedy that remains available to the aggrieved party is only recourse to a court of law. However, in the reverse scenario that the platform dismisses the complaint and refuses to make the content available again, it must offer out-of-court dispute settlement to the author.[###] The body conducting the out-of-court proceedings must meet basic requirements of neutrality, fairness and expertise. These are familiar from the European Directive on alternative dispute resolution for consumer disputes.[§§§§] If the ADR entity finds in favour of the party who has lodged the complaint, i.e. the author of the incriminated content, and orders the platform to reinstate such content, the platform is bound by the decision.[*****]

[###] Art. 18 DSA

[§§§§] Art. 6–9 Directive 2013/11/EU 3.

[*****] In the reverse scenario that the ADR entity finds in favour of the platform by rejecting the complaint, the author is not bound by the decision but free to seek recourse in court.

It is striking that this DSA-Proposal has an asymmetric design. The complaint-handling mechanism and the ADR-Procedure are only open to the author of the content or the owner of the account, respectively, and not to the party which is affected by a decision *not* to remove the incriminated content. An aggrieved party, e.g. a victim of defamatory statements, cannot enter the complaint-handling mechanism. It can only file a suit in a court of law and as we all know, a reputation may be lost on the internet long before a public court even issues an interim order of protection. Obviously, the DSA mechanisms are designed to prevent overblocking but do not care about underblocking. The redress-mechanisms ensures control of any removal of content but is not open for re-evaluation of any decision against removal. This is an imbalance which raises constitutional concerns.

The decision of the platform whether or not a posting has violated the personality rights of another party depends on the application of legal norms which strike a balance between the author's freedom of expression and the third party's right to privacy. The decision whether to remove requires the weighing and balancing of conflicting fundamental rights which are enshrined in national constitutional law as well as in European law. The procedural remedies provided by the DSA should reflect this substantive duty and should provide access for both sides, the aggrieved party as well as the user who has generated content.

The German Network Enforcement Act requires only a complaint-handling system, but this system is open to both sides of the conflict.

## V. Private Content Moderation: Regulatory Reluctance

### 1. Private content Moderation and the scope of procedural requirements

The Notice and take-down mechanism first of all is a legal remedy against infringements of private rights or other unlawful content. The German Network Enforcement Act as the first legislation in the field, had limited its procedural and organizational requirements on notices about specific unlawful acts. However, Social Media Platforms have operated some kind of notice-and take down procedures for violations of community standards independent from such regulatory frameworks for a long time. As legal limitations of free speech and community standards overlap in many respects, the field of application of legally required procedures and such voluntary arrangements is also overlapping. The evaluation of the German Network Enforcement Act has demonstrated that parallel procedures entail difficult regulatory delineations and it is obvious that users should have some easily accessible remedies not only against removal of content which is based on a contested violation of law but also against removal of content which is based on a contested violation of community standards. Therefore, the amendment of the NEA and the proposed DSA rightly apply the just described complaint-handling and out-of-court settlement mechanisms to all kinds of decision on the removal of content irrespective of its legal basis. In other words, any removal based on the violation of community standards is also to be re-evaluated upon request. This provides safeguards against an abusive application of community standards by platforms.

Additionally, the proposed DSA requires platforms to include information on any restrictions that they impose on user-generated content and any policies, procedures, measures and tools used for content moderation in their terms and conditions and to make the information also available to the public.[††††]

However, there are no regulations at all concerning the substance of such standards. The review-mechanisms cannot accomplish more than a proper application of given standards. The crucial question concerns the legitimate substance of such standards. Do they constraint public

[††††] Art. 5 DSA

discourse in an inappropriate manner? The DSA proposal promises in Art. 1 to effectively protect the fundamental rights enshrined in the Charter. Is it an empty promise with regard to free speech?

## 2. Substance of Community Standards beyond Regulation yet Critical

Let us start by analyzing the incentive structure. A Social Media platform which is ultimately selling custom-tailored attention to the advertising industry, strives to maximize the number of users who are as active as long as possible in an online community, and community standards as well as content moderation and curation are used to foster an encouraging online environment. Standards will therefore depend on the attracted communities, they may prefer some kind of attention-grabbing content and exclude some kind of controversy-prone content as it has been analysed for content-moderation tools. In any event, they will probably differ from the scope of lawful speech.

In principle, it is without doubt the right of any private provider of a service to define the terms and conditions under which the service can be used. This would not raise many concerns if there were plenty of platforms so that each user could choose the terms and condition he or she preferred. However, as we all know, there aren´t many platforms. Economies of scale and network-effects were strong incentives and drivers for fast growth and have resulted in only a few social Media platforms dominating the market. Additionally, strong lock-in effects for the users hamper the contestability of the markets, making it difficult for any company to enter it. As a consequence, market power transforms into the power to limit what can be said and heard in increasingly important spheres of public discourse. The "de-platforming" of individuals (e.g. former US-President Trump) on grounds of violation of community standards is the most obvious example. In sum, the DSA promise is not self-fulfilling and regulatory reluctance is questionable.

If we think about regulation we should bear in mind that we have to strike the balance between the fundamental right of the platform operators to conduct their business as they please on the one hand and the rights to non-discrimination and to freedom of expression including the public interest in an open public discourse on the other hand. Any regulation needs to be graded and sensitive to the respective importance of the platform for public discourse. Broadly speaking, operators should be able to use community standards as a means of product differentiation, but: they should be prevented from discriminating users or content without legitimate reason. They should also be required to define the community standards the more according to the general laws the more the service aims at becoming the overall online environment for its users and the more important it becomes for public discourse. Facebook probably comes to mind when we look for an example for a network which is a particularly walled-garden with great impact on public discourse.

An appropriate regulation will rather result in a broad clause which requires operators to draft the terms and conditions with due regard to the rights and legitimate interests of its users than in precise conditions. This, however, is very familiar from traditional delineations between conflicting private rights and has become the characteristic feature inter alia of data protection law.

## 3. Human Rights Influence on Private Governance?

Even if such statutory requirements seem to be desirable, thus far, there aren´t any. For the moment, the only source for any influence of fundamental rights on the drafting of community standards can be the constitution and the European Charter, respectively. This touches upon the tricky question of horizontal effects of fundamental rights on private transactions and, more

generally, on the acts of private parties.#### While we have a long standing doctrine on indirect effects of fundamental rights in private law at the national level in Germany, there is no broadly accepted doctrinal basis for extending the scope of the EU-Charter to relations between private parties.§§§§§ Thus, its effect on the process of drafting terms and conditions that specify the requirements for removal and reinstatement of undesirable content remains an open question.

It is plausible to assume, however, that the Union has a duty to protect all fundamental rights when setting rules for the online environment. And if the online environment is shaped by powerful private actors, it is also plausible to assume that the duty to protect may comprise making those private actors responsive to the fundamental rights that their business practices may affect. It is at least doubtful, whether the broad discretion that the DSA leaves to online-platforms with respect to their terms and conditions is consistent with the Charter.****** Arguably, it is not just desirable but demanded to enact a statutory requirement which forces operators to draft the community standards with due regard to users´ right to freedom of expression.

## VI. Systemic Effects: Uncertainty

We have now touched liability for illegal content and legal restrictions for community standards.

The most difficult regulatory question, however, is definitely how to deal with systemic effects.

There are two major reasons for it. The first reason is that we have thus far very little knowledge about such systemic effects. The general notion that there are toxic effects for public discourse and an erosion of trust in public institutions is based on shocking events like the storming of the United States Capitol. However, it is not easy to distinguish between Social Media impact on the one hand and developments that have an interplay with it but are rooted outside the internet on the other hand. Is polarization resulting from Social Media Communications or is Social Media Communications just a mirror?

The second reason is that any regulation would go beyond any traditional limitations of free speech and would entail a great potential for abuse. The right to freedom of expression protects for good reasons speech which is annoying or irritating and viewpoint-neutrality is a general rule for limitations.

### 1. Protecting liberal Democracy and Danger of Abuse

The high degree of uncertainty concerning systemic risks has become a major driver of the quest for more transparency and further risk assessment. The DSA approach relies on knowledge generation and accountable self-regulation. The DSA imposes a duty on very large platforms to

---

####Cf. Kühling, "Fundamental Rights", in von Bogdandy and Bast, Principles of European Constitutional Law, 2nd ed. (C.H.Beck, 2009), pp. 479–515, at pp. 492–494; with regard to Art. 10 ECHR (in an extreme case) see also ECtHR, Dink v. Turkey, Appl. No. 2668/07 et al., chamber judgment of 14 September 2010, with broad wording in para 106: "…que l'exercice réel et effectif de la liberté d'expression ne dépend pas simplement du devoir de l'Etat de s'abstenir de toute ingérence, mais peut exiger des mesures positives de protection jusque dans les relations des individus entre eux. En effet, dans certains cas, l'Etat a l'obligation positive de protéger le droit à la liberté d'expression contre des atteintes provenant même de personnes privées…".

§§§§§ According to Art. 51 of the Charter, its scope is confined to acts of the state, as defined in this article. Even though the fundamental freedoms of the TFEU and Art. 21 of the Charter (non-discrimination) have been applied to private parties, there is no general doctrine on horizontal effects.

****** This is especially true considering the potential consequences of Art. 12 DSA. If it were interpreted to implicitly grant the platforms freedom to enact terms and conditions of their choice, unrestrained by the fundamental rights guaranteed in the Charter, Art. 12 DSA would even foreclose judicial review at the national level. This would result in a substantial decrease in the level of fundamental rights protection which would be at odds with the ambitions of Art. 1 (2) (b) DSA and can hardly be consistent with the true intentions of the Commission. In order to bring the DSA in line with the Charter, it is necessary to include in the DSA a provision that explicitly authorizes the Member States to review platform terms and conditions in light of the fundamental rights enshrined in the Charter or national constitutional law.

identify, analyse and assess periodically any significant systemic risk stemming from the functioning and use made of their services in the EU (Art. 26 DSA). Following up the risk-assessment, the platforms have to put in place mitigation measures and to publish a report on risk-assessment and measures taken. Besides, public authorities can request very large platforms to provide access to data to vetted researchers (Art. 31 (2) DSA). Eventually, if significant systemic risks emerge and concern several large platforms, the Commission may invite the very large platform and other stakeholders to draw up a code of conduct (Art. 35 (2) DSA).

### 2. Regulating Counter speech, Dynamics and Persistence

Given the current state of uncertainty, this might be the best approach for the moment. In the long run, however, it can hardly be left to the platforms alone to define the degree of bearable risks and appropriate countermeasures. If we think about adequate means, however, they should be as transparent as possible and extremely cautious with regard to removal of content. Even though it is too early to determine any measure, three approaches appear to be at least plausible. Regulation could foster counter-speech for some kinds of content such as Fake-News, it could slow down the specific dynamics of Social Media Communications if pattern occur which proved to typically attack vulnerable groups or it could limit persistence of postings in order to avoid a harmful amount of toxic speech. All such approaches, however, need more insights and should be tailored very carefully.

## VII. Concluding Remarks

Systemic effects deserve very much attention in the future but this lecture has to end pretty soon. Instead of elaborating about the upcoming questions, let me make some concluding remarks.

Responsibility of Social Media is highly ranked on the international agenda. Apart from an obvious abuse of the situation by authoritarian regimes, we observe a serious discussion among liberal democracies about future regulatory frameworks. All liberal democracies find themselves in a dilemma. They are based on free speech and yet find themselves increasingly threatened by speech. Government must not abridge free speech but faces powerful platforms and private speech regulation which may put open public discourse at risk. The emerging European way to deal with the dilemma is a graded system of Co-Regulation. Government imposes procedural and organizational obligations on the Social networks which ensure prompt removal of illegal content and proper evaluation of the underlying decision on the content. It will require an application of community standards which is sensitive to fundamental rights and arguably, it will also have to provide for an influence of fundamental rights on the substance of community standards. Finally, it aims at shedding more light on systemic effects by fostering risk-assessment and independent research. To my mind, this is not bad for a beginning. I am curious to get to know your perspective in the following discussion. Thank you very much for your attention.