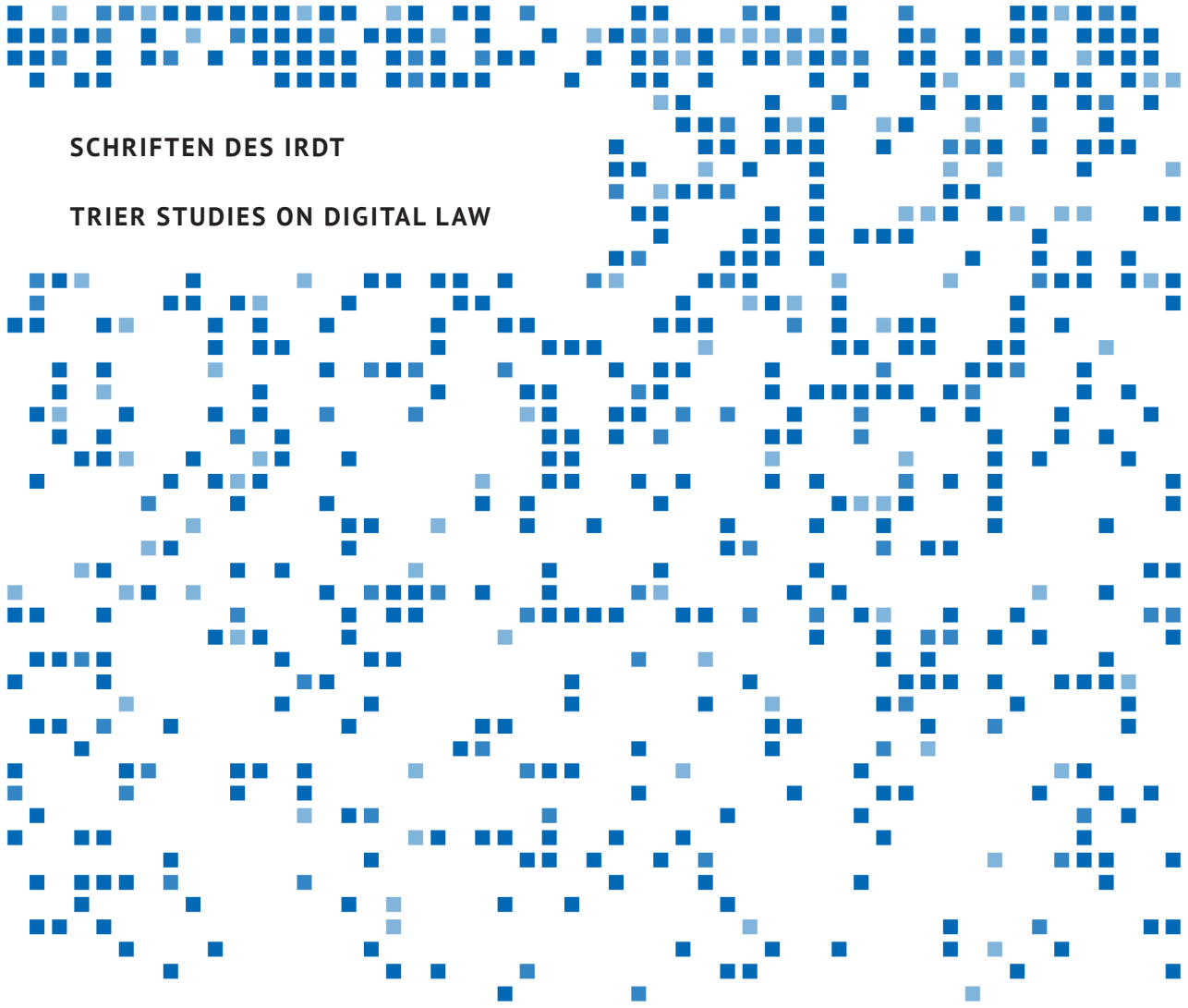

SCHRIFTEN DES IRDT

TRIER STUDIES ON DIGITAL LAW

Antje von Ungern-Sternberg (ed.)

Content Regulation in the European Union

The Digital Services Act

Volume 1

Antje von Ungern-Sternberg (ed.)

Content Regulation in the European Union

The Digital Services Act

TRIER STUDIES ON DIGITAL LAW

Published by: Verein für Recht und Digitalisierung e.V.
Institute for Digital Law Trier (IRDT)

Trier, 2023
Volume 1

SCHRIFTEN DES IRDT

TRIER STUDIES ON DIGITAL LAW

Published by Verein für Recht und Digitalisierung e.V.,
Institute for Digital Law Trier (IRDT), Behringstr. 21, 54296 Trier, Deutschland

The German National Library lists this publication in the German National Bibliography; detailed bibliographic information is available at http://dnb.d-nb.de
This book is also available as E-Book at https://www.epubli.com/shop

ISBN: 9783757550950
URN: urn:nbn:de:hbz: 385-2023051001
DOI: https://doi.org/10.25353/ubtr-xxxx-3a52-23eb

The Trier Studies on Digital Law are sponsored by Trier University and the IRDT.

UNIVERSITÄT TRIER

IRDT Institute for Digital Law Trier

Foreword

Good books deserve to be read. This is especially true in the digital age. Precisely due to the easy availability of information in legal databases, there is a danger that readers will no longer seek out the printed book. This is all the more true if obtaining it involves long trips to the library, waiting times for orders, and financial costs. But those who only publish on the internet are often not perceived as relevant or must pay high Open Access fees.

With the **Trier Studies on Digital Law**, we, therefore, want to contribute to our goal of a society-oriented science that is as freely accessible as possible. By way of Gold Open Access, we offer with the series a platform to share scholarly writings in the field of digital law with a broad audience.

The Trier Studies on Digital Law are characterised by four guiding principles:

- ***Relevance & Quality*:** The books we include in our series are excellent works on relevant topics in the law of digitalisation.
- ***Free Access Culture*:** For readers, publications are freely available in digital form, at no cost. The publication is also free of charge for authors.
- ***Fast & Permanent Availability*:** Publications are available on the internet shortly after their completion. Cooperation with Trier University Library and the German National Library guarantees that this remains the case permanently.
- ***Digital & Analogue*:** The digital version of a publication can be accessed from anywhere at the touch of a button. Those who would like to hold the publication in their hands or study it in-depth can also purchase physical copies by way of print-on-demand services.

With our publication model, we thus assure compliance with quality standards and enable both authors and readers to participate in the scientific discourse without having to make considerable financial contributions or resort to public funding.

The Trier Studies on Digital Law are published by the **Verein für Recht und Digitalisierung e.V.**, a non-profit association dedicated to support the IRDT, the Institute for Digital Law Trier.

Peter Reiff
Chair, Verein für Recht und Digitalisierung e.V.

Content Regulation in the European Union: Hate Speech, Fake News & Co – An Introduction

Digital platforms like Facebook, YouTube, TikTok and Twitter play a key role in both private and public communication. The platforms enable everyone to contribute to matters of public debate without the need to pass classic gatekeepers such as newspapers, radio or TV stations. This realises the idea of a very large and open marketplace of ideas, which is at the heart of democracy. Gradually, however, and increasingly in recent years, the dark side of unfiltered online content has become visible: Hate speech mobilises against minorities, incites crime and intimidates politicians. Fake news threaten the fairness of electoral campaigns and the integrity of election results, they help spread conspiracy theories and prevent an effective fight against the pandemic. Other forms of problematic content include child pornography, violent videos, copyright infringement, or leaks of private data.

Up to this point, digital platforms have resorted to self-regulation based on community standards or other forms of private standard setting in order to draw the fine line between forbidden content and free speech. Thus, private companies effectively define the scope and the limits of freedom of opinion within their digital spheres. However, European states and the European Union have also started legislating, as the German Network Enforcement Act (2017), the French Law on the fight against the manipulation of information (2018), and the EU Regulation on the dissemination of terrorist content online (2021) show. In addition to sector-specific rules, the EU's Digital Services Act, the DSA (2022), introduces an updated horizontal framework for all categories of content, products, services and activities on intermediary services and aims to harmonise the fragmented rules of platform liability. This raises several questions. How far do – national and European – free speech guarantees go? If hate speech and defamation can be banned to protect the victims' rights, how can the prohibition of fake news be justified? What is the remaining leeway of the platforms for private content moderation? Who is responsible for fighting and taking down illegal content? How can the victims of de-platforming, content takedowns or shadow banning claim their right to freedom of opinion? Finally, how will these legal responsibilities be enforced?

These questions were discussed at the annual conference of the IRDT, the Digital Law Institute Trier, on 13 and 14 October 2022, at the Electoral Palace in Trier. The written contributions are collected in this edited volume. The volume opens with *Katharina Kumkar's* introduction and overview of the Digital Services Act as it was finalised in October 2022. This legal introduction is complemented by a technical one. *Martin Steinebach* gives a technical assessment of potentials and limits of filter technology for the regulation of hate speech and fake news. The following two contributions

take a European perspective. *Antje von Ungern-Sternberg* analyses how the challenges of content moderation are addressed by EU legislative activities (including the DSA) and lead to a Europeanisation of Freedom of Expression. *Mattias Wendel* considers the pluralist structure of the European fundamental rights architecture and its interplay with EU legislation as a matter of "Taking or Escaping Legislative Responsibility?". The final contributions provide an in-depth assessment of specific aspects of the Digital Services Act. *Florence G'sell* examines its crucial provisions and focusses particularly on possible challenges regarding the DSA's implementation. *Ruth Janal's* article is dedicated to the platforms' community standards restricting lawful but harmful content, and contrasts the German Federal Court of Justice's approach and the legal situation under the DSA.

Antje von Ungern-Sternberg
Editor of the Volume

Content

The Digital Services Act: Introduction and Overview

Lea Katharina Kumkar

As of 16 November 2022, the Digital Services Act (DSA)[1] has come into force. Along with the Digital Markets Act (DMA),[2] which deals with the competitive aspects of large online platforms, the DSA is part of the legislative package proposed in December 2020 by the European Commission to upgrade rules governing digital services in the EU. Together they form a single set of new rules that are applicable across the whole EU, intended to create a safer and more open digital space. The two main goals of these pieces of legislation are to create a safer digital space in which the fundamental rights of all users of digital services are protected and to establish a level playing field to foster innovation, growth, and competitiveness, both in the European Single Market and globally.[3]

I. Background

Until now, the liability of online service providers has been largely determined by the so-called E-Commerce Directive (Directive 2000/31/EC[4]), which dates to the turn of

[1] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC, the Digital Services Act, 2022, OJ/L 277/1. For an overview, see also Chiarella 'Digital Markets Act (DMA) and Digital Services Act (DSA): New Rules for the EU Digital Environment', 2022, 9 Athens Journal of Law 1, 11 et seq.; Gerdemann and Spindler 'Das Gesetz über digitale Dienste (Digitals Services Act)', 2023, GRUR 3 (part 1) and 115 (part 2); Raue and Heesen 'Der Digital Services Act', 2022, NJW 3357; Wilman 'The Digital Services Act (DSA) – An Overview', 16 December 2022, https://ssrn.com/abstract=4304586.

[2] Regulation EU 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives EU 2019/1937 and EU 2020/1828 (the Digital Markets Act), 2022 O.J. (L 265), 1. For an overview, see, e.g., Chiarella 'Digital Markets Act (DMA) and Digital Services Act (DSA): New Rules for the EU Digital Environment', 2022, 9 Athens Journal of Law 1, 5 et seq.; de Streel and Alexiadis 'The EU's Digital Markets Act – Opportunities and Challenges Ahead', 8 June 2022, https://ssrn.com/abstract=4131781; Kumkar, 'Neue Impulse für den Wettbewerb auf digitalen Märkten', 2022, RDi 347; van den Boom 'What does the Digital Markets Act harmonize? – Exploring interactions between the DMA and national competition laws', 2022, ECJ, https://doi.org/10.1080/17441056.2022.2156728.

[3] See European Commission https://digital-strategy.ec.europa.eu/de/policies/digital-services-act-package.

[4] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 'Directive on electronic commerce'.

the millennium. However, the digital landscape has changed considerably since the directive was issued over 20 years ago, and many of its provisions are no longer up-to-date in view of new business models and changing user habits. Digital platforms such as Google, Amazon, Zalando, Facebook, Instagram, TikTok, and YouTube have gained enormous importance. These online platforms have created significant benefits for consumers and for innovation. They have helped to make the European Union's internal market more efficient and to facilitate cross-border trade within and outside the Union. However, these platforms have also come under increasing criticism in recent years. One major problem is the trade and exchange of illegal goods, services, and content on the Internet. Furthermore, online services are more and more frequently misused by manipulative algorithmic systems to increase the spread of disinformation and to cause harm in other ways. These new challenges and the way platforms address them are considered to have significant implications on fundamental rights, which is why the European legislature saw a need for immediate action.[5]

Against this backdrop, the DSA aims to modernize the legal framework of the E-Commerce Directive and implement harmonised binding obligations for digital services throughout the European Union. The DSA explicitly restates this objective in Art. 1 para 1 DSA: "*The aim of this Regulation is to contribute to the proper functioning of the internal market for intermediary services by setting out harmonised rules for a safe, predictable and trusted online environment that facilitates innovation and in which fundamental rights enshrined in the Charter, including the principle of consumer protection, are effectively protected.*" By choice of the legal form of a regulation, the European legislature intends to ensure the effectiveness of the rules laid down in the DSA and to create a level playing field within the internal market (cf. DSA, recital 7).

Just like the concurrently drafted Digital Markets Act,[6] the DSA has gone through the legislative process in a rush. The Commission proposed it in December 2020, and on 23 April 2022 – a mere 16 months later – a political agreement between the EU institutions was reached on a final version of the text as part of the so-called trilogue process. The DSA was published in the Official Journal on 27 October 2022; the legislation entered into force on 16 November 2022 and will apply in all EU countries as of 17 February 2024. The pace at which the legislation progressed is even more impressive considering the development of the provisions themselves. After all, the political bodies involved were by no means in agreement on all points. On the contrary, during the legislative process, the DSA was the subject of a highly controversial debate, with almost

[5] See European Commission https://digital-strategy.ec.europa.eu/de/policies/digital-services-act-package.

[6] On 25 March 2022, a political agreement was reached on the DMA, which was published in the Official Journal on 12 October 2022. On 1 November 2022 (20 days after the publication in the Official Journal), the DMA entered into force and will apply from 2 May 2023.

three thousand amendments submitted by the EU bodies involved. Some of the most controversial issues included the additional provisions for very large search engines (the so-called VLOS, see also below IV.4), compensation rights for users, and rules for algorithm-based recommender systems, target group advertising, and dark patterns.[7]

II. Scope of Application and Regulatory Structure of the DSA

The scope of application is laid down in Art. 2 DSA. According to Art. 2 para 1 DSA, the application is limited to "intermediary services", which are – as Art. 3 lit. g DSA specifies in more detail – all information society services that transmit or store user information at the request of recipients; this essentially includes all kinds of host and access providers. Regarding the territorial scope, the DSA follows the so-called "marketplace principle" known from the GDPR.[8] According to Art. 2 para 1, the DSA applies to all intermediary services offered to recipients within the European Union, irrespective of where the providers of those intermediary services are located.[9]

In addition to the liability rules, the DSA introduces a whole new set of specific due diligence requirements for intermediary service providers. Combined, these rules intend to set new standards for the accountability of online platforms regarding illegal and harmful content. Following the general provisions in the first chapter, the DSA can thus be divided into three key sections: The liability rules for intermediaries in Chapter II (see below III), the due diligence requirements in Chapter III (see below IV), and the implementation and enforcement provisions in Chapter IV (see below V). The regulations within the DSA are designed as a "pyramid model",[10] i.e., in addition to some provisions for all intermediary services, certain regulations only apply to services of a certain size (VLOPs and VLOS, see below IV.4) or from certain sectors (e.g., hosting services or online marketplaces, see below IV.2 and IV.3).

[7] Cf. Gerpott 'Das Gesetz über digitale Dienste nach den Trilog-Verhandlungen', CR 2022, 516, 519 et. seq.

[8] Cf. Art. 3 GDPR.

[9] This represents a change compared to the E-Commerce Directive, which was still based on the 'country of origin principle' (cf. Art. 3 the E-Commerce Directive).

[10] See Spindler 'Der Vorschlag für ein neues Haftungsregime für Internetprovider – der EU-Digital Services Act (Teil 1)', GRUR 2021, 545 ("Pyramiden-Modell").

III. Liability Rules for Providers of Intermediary Services (Chapter II)

Chapter II, which comprises Art. 4 to 10 DSA, contains the liability rules for intermediary services previously included in Art. 12 to 15 of the E-Commerce Directive. The implementation of an appropriate liability model for intermediary services is proving to be a balancing act between responsibility and privileging.[11] On the one hand, it is not justifiable to completely exempt intermediary services from liability. After all, many of these services create a potential source of harm to others (e.g., by enabling infringement of third parties' rights on a large scale), mainly for their own profit. On the other hand, intermediary services cannot be held responsible for any and every infringement either. This is because the Internet is based on the possibility of unhindered exchange of information and is thus inextricably linked to the business models of information intermediaries.[12] These business models would be threatened by a liability model that is too rigid: In the worst case, the fear of subsequent lawsuits would push intermediary services into the role of censorship authorities, and user contributions would be deleted in cases of doubt for fear of the services' own liability, which could finally result in overblocking.[13] This is the reason neither extreme, no liability nor full liability, is fully desirable. Rather, to ensure that Europe's digital economy remains viable while effectively protecting European values such as freedom of expression, democracy, and diversity, the two extremes must be balanced. This rationale was already inherent in the former liability regime in Art. 12 to 15 E-Commerce Directive, whose framework has been largely adopted by the DSA.[14] However, some adjustments and additions have been made to adequately reflect the changed role and importance of digital intermediaries.

As under the E-Commerce-Directive, the DSA merely defines the limits of intermediary liability. Liability itself is governed – as before – by national legislation. In the case of pure intermediary services – that is, "mere conduit", "caching", and "hosting" – there is generally no responsibility for third-party content under the DSA unless the intermediary does have actual knowledge or becomes aware of illegal activity or illegal content (the so-called notice-and-take-down procedure, cf. Art. 3 to 6 DSA). However,

[11] See Hofmann 'Die Haftung von Internetplattformen und Internetzugangsvermittlern – Auf dem Weg von der Sonderdogmatik zum Kernzivilrecht', ZfPW 2021, 385, 386.

[12] Cf. Hofmann 'Die Haftung von Internetplattformen und Internetzugangsvermittlern – Auf dem Weg von der Sonderdogmatik zum Kernzivilrecht' ZfPW 2021, 385, 386; Janal 'Haftung und Verantwortung im Entwurf des Digital Services Acts', 2021, ZEuP 227, 237 et seq.

[13] See Kumkar 'Die Plattform-Verantwortlichkeit für Schwarmschäden' in Köhler and Korch (eds.) 'Schwäme im Recht', 2023, p. 108.

[14] See Cauffman and Goanta 'A New Order: The Digital Services Act and Consumer Protection', 2021, 12 EJRR 758, 763.

the DSA recitals stipulate that these privileges "*should not apply where, instead of confining itself to providing the services neutrally by a merely technical and automatic processing of the information provided by the recipient of the service, the provider of intermediary services plays an active role of such a kind as to give it knowledge of, or control over, that information*"[15] or "*[w]here a provider of intermediary services deliberately collaborates with a recipient of the services in order to undertake illegal activities*".[16] Hence, as in the context of the E-Commerce Directive, under the DSA, preferential treatment is therefore dependent on *neutral behavior*. As with Art. 15 E-Commerce-Directive, the DSA prohibits general monitoring obligations: No general obligation shall be imposed on providers to monitor the information which they transmit or store or to actively seek facts or circumstances indicating illegal activity (cf. Art. 8 DSA).

For host providers (only), this "notice-and-take-down" approach is evolving into a "notice-and-action" mechanism: Art. 16 DSA stipulates that providers of hosting services need to put mechanisms in place to allow users to notify them of the presence of specific items of information that are considered to be illegal content.[17] Those mechanisms must be easy to access and user-friendly and allow for the submission of notices exclusively by electronic means (cf. Art. 16 para 1 DSA). These notices are then considered to give rise to actual knowledge or awareness in respect to the specific item of information concerned, thereby incurring liability under Art. 6 DSA, to the extent that they allow for identifying the illegality of the relevant activity or information without a detailed legal examination (cf. Art. 16 para 3 DSA). In addition, host providers must comprehensively inform those affected by the notification of illegal content about the notification itself, any blocking and deletion decisions, and if applicable, the use of automated systems (cf. Art. 16 par. 4 to 6 DSA).

New additions to the existing legal framework can be found primarily in Art. 7 to 10 DSA. One important new feature is Art. 7 DSA, according to which providers shall not be deemed ineligible for the exemptions from liability solely because they carry out voluntary own-initiative investigations or take other measures aimed at detecting, identifying, and removing or disabling access to illegal content, meaning that a purely voluntary review of user content shall no longer lead to the exclusion of liability privileges

[15] Cf. DSA, recital 18.

[16] Cf. DSA, recital 20. However, it remains unclear if the neutrality requirement is intended to be applicable in the same way as it was understood in the underlying case law, cf. Cauffman and Goanta 'A New Order: The Digital Services Act and Consumer Protection' 2021, 12 EJRR 758, 764 et seq.

[17] For the problems in evaluating the illegality of content, see Korpisaari 'From Delfi to Sanchez – when can an online communication platform be responsible for third-party comments? An analysis of the practice of the ECtHR and some reflections on the digital services act', 2022, J. Media Law 1, 23 et seq.

(the so-called Good Samaritan Rule).[18] By taking this approach, the regulation addresses the potential risk that service providers will refrain from taking proactive measures to protect against infringements for fear of leaving the realm of "neutrality".[19] In addition, two new provisions are included in Art. 9 and 10 DSA, which refer to official orders issued by the relevant national judicial or administrative authorities to act against illegal content or to provide specific information about individual users of their service.[20]

IV. Due Diligence Requirements for a Transparent and Secure Online Environment (Chapter III)

The actual core of the DSA lies in the third chapter. Here, in a total of six subsections, the regulation addresses due diligence requirements for a transparent and secure online environment. These requirements vary depending on the size and nature of the online services, reflecting the DSA's very own "pyramid-model" (see II above).

1. Provisions Applicable to All Intermediary Services

For all intermediary services, the DSA provides for a number of information and transparency obligations, including the establishment of points of contact (cf. Art. 11 and 12 DSA) and the designation of legal representatives if there is no establishment in the European Union (cf. Art. 13 DSA), as well as annual transparency reports (cf. Art. 15 DSA).[21] Of considerable significance is the requirement that Art. 14 DSA imposes on the intermediary services terms and conditions. Not only must the terms and conditions of the platform meet certain transparency requirements (cf. Art. 14 para 1 to 3, 5 and 6 DSA), the DSA further requires that when applying or enforcing any kind of restrictions in relation to the use of their services, the providers of intermediary services shall act in "*a diligent, objective and proportionate manner [...] with due regard to the rights and legitimate interests of all parties involved, including the fundamental rights*

[18] See also Korpisaari 'From Delfi to Sanchez – When can an online communication platform be responsible for third-party comments? An analysis of the practice of the ECtHR and some reflections on the digital services act', 2022, J. Media Law 1, 25.

[19] See Janal 'Haftung und Verantwortung im Entwurf des Digital Services Acts', ZEuP 2021, 227, 238 et seq.

[20] See Chiarella 'Digital Markets Act (DMA) and Digital Services Act (DSA): New Rules for the EU Digital Environment', 2022, 9 Athens Journal of Law 1, 13.

[21] See Spindler 'Der Vorschlag für ein neues Haftungsregime für Internetprovider – der EU-Digital Services Act (Teil 1)', GRUR 2021, 545, 551 et seq.

of the recipients of the service", effectively committing intermediary services to the rights enshrined in the European Charter of Fundamental Rights (cf. Art. 14 para 4 DSA).

Also applicable to all online services, regardless of their size and nature, are the provisions on self-regulatory measures as set out in Chapter II Section 6 (Art. 44 to 48 DSA). This includes the development of voluntary standards (Art. 44 DSA) and different kinds of codes of conduct (Art. 45 to 47 DSA), as well as crisis protocols (Art. 48 DSA).[22]

2. Additional Provisions for Hosting Services

According to Art. 16 et seq. DSA, additional obligations are imposed on the providers of hosting services, including online platforms. Hosting services are intermediary services that store the information provided by, and at the request of, a recipient of the service (see Art. 2 lit. g DSA); this includes, for example, search engines, social networks, cloud computing, or web hosting services. In addition to the aforementioned notice-and-action mechanism (Art. 16 DSA), further content management obligations are imposed on hosting services. These include the obligation to provide a clear and specific statement of reasons to users affected by restrictions imposed on the grounds that the information is illegal content or incompatible with the terms and conditions of the service (Art. 17 DSA), as well as a notification obligation in the event of suspected criminal offences: When a provider of hosting services becomes aware of any information giving rise to a suspicion of a criminal offence involving a threat to the life or safety of a person, it shall promptly inform the law enforcement or judicial authorities of the Member State or Member States concerned in its suspicion and provide all relevant information available (cf. Art. 18 DSA).

3. Additional Provisions for Online Platforms and Online Marketplaces

Online platforms are subject to several additional provisions set out in Art. 19 to 28 DSA. However, micro and small enterprises are largely exempted from these obligations (cf. Art. 19 DSA). In line with the terminology of the DSA, "online platforms" are a special form of hosting service, whose characteristic feature is that they not only store users' information but also disseminate it to the public at the request of a recipient of the service (cf. Art. 2 lit. i DSA). According to Recital 13 of the DSA, this definition includes social networks such as Facebook, YouTube, and TikTok as well as online platforms "allowing consumers to conclude distance contracts with traders", i.e.,

[22] See also Cauffman and Goanta 'A New Order: The Digital Services Act and Consumer Protection', 2021, 12 EJRR 758, 767 et seq. ('Outsorcing solutions to private entities').

online-marketplaces such as eBay or Amazon Marketplace. Online platforms are required to set up internal complaint-handling systems (Art. 20 DSA), procedures for out-of-court dispute settlements (Art. 21 DSA), and mechanisms to ensure that notices about illegal content submitted by so-called "trusted flaggers" are prioritized (Art. 22 DSA).[23] According to Art. 23 DSA, online platforms are obliged to take action against misuse (e.g., suspension of users that frequently provide manifestly illegal content or that frequently submit notices or complaints that are manifestly unfounded, cf. Art. 23 para 1 and 2 DSA). Further, the DSA provides some additional transparency reporting obligations for providers of online platforms (Art. 24 DSA); special requirements for the design and organization of online interfaces, advertising, and recommender systems (Art. 25 to 27 DSA); and a specific provision dedicated to the protection of minors in the online environment (Art. 28 DSA).

In Articles 29 to 32, the DSA provides for additional provisions applicable to the subcategory of "online platforms allowing consumers to conclude distance contracts with traders", i.e., online marketplaces. Again, micro and small enterprises are largely exempt from these obligations (cf. Art. 29 DSA). Art. 30 DSA is intended to ensure the "traceability of traders". Thus, providers are obliged to admit only those traders to their platform about whom they have certain information (cf. Art. 30 para 1 DSA). If necessary, the provider shall endeavor to verify the reliability of the information provided (cf. Art. 30 para 2 DSA).[24] In addition, the DSA implements specific requirements for online marketplaces for the design of their online interfaces (Art. 31 DSA) as well as information obligations towards consumers who have purchased illegal products or services through the services of the platforms (Art. 32 DSA).[25]

4. Additional Obligations for Very Large Online Platforms and Very Large Search Engines (VLOPs and VLOS)

The requirements in Section 5 of Chapter II (Art. 33 to 43 DSA) apply only to the very large online platforms and online search engines with an average number equal to or higher than 45 million monthly active users in the European Union (cf. Art. 33 para 1 DSA), the so-called VLOPs (Very Large Online Platforms) and VLOS (Very Large Online Search engines). Given their critical importance for the exchange of public opinion and online commerce,[26] the DSA imposes additional compliance obligations on

[23] See also Chiarella 'Digital Markets Act (DMA) and Digital Services Act (DSA): New Rules for the EU Digital Environment', 2022, 9 Athens Journal of Law 1, 14 et seq.

[24] For this traceability obligation, see in detail Rott 'New Liability of Online Marketplaces Under the Digital Services Act?', 2022, 6 Eur. Rev. Priv. Law. 1, 10 et seq.

[25] See also Rott 'New Liability of Online Marketplaces Under the Digital Services Act?', 2022, 6 Eur. Rev. Priv. Law. 1, 13 et seq.

[26] See DSA, recitals 75 and 76.

these very large platforms. However, and this is a significant difference from the obligations outlined in Sections 1 through 4, these special provisions apply only if the platforms have been designated by the European Commission as VLOPs or VLOS, respectively (see Art. 33 para 1 DSA: "*This Section shall apply to online platforms and online search engines [...] which are designated as very large online platforms or very large online search engines pursuant to paragraph 4.*"). The specific details of the designation process are set out in Art. 33 para 4 to 6 DSA. If during an uninterrupted period of one year, the number of average monthly active users falls below the 45 million threshold referred to in Art. 33 para 1, the Commission shall terminate the designation (cf. Art. 33 para 5 DSA).

The VLOPs and VLOS face some particularly far-reaching obligations. For example, providers of VLOPs and VLOS must conduct risk assessments that "*diligently identify, analyse and assess systemic risks stemming from their service and its related systems*"; this includes the risks of dissemination of illegal content as well as any actual or foreseeable negative effects on fundamental rights, on civic discourse and electoral processes, public security, gender based violence, the protection of public health and minors, or serious negative consequences to the person's physical and mental well-being (cf. Art. 34 DSA).[27] In order to mitigate these risks, pursuant to Art. 35 DSA, providers of VLOPs and VLOS are to put in place suitable countermeasures that are "*tailored to the specific systemic risks identified pursuant to Article 34, with particular consideration to the impacts of such measures on fundamental rights*", e.g., by adapting the design, features, or functioning of their services; by changing their terms and conditions; or by taking awareness-raising measures (see full list in Art. 35 para 1 DSA). Furthermore, Art. 36 DSA implements a special "crisis response mechanism", under which VLOPs and VLOS may be required by the European Commission to take special measures in case of crisis (see full list of possible actions in Art. 36 para 1 DSA). In addition, the DSA provides for the implementation of a monitoring system for VLOPs and VLOS, both externally (see Art. 37 DSA) and internally (see Art. 41 DSA), and introduces some additional transparency obligations (see Art. 39 and 42 DSA). Art. 40 DSA grants the Digital Services Coordinator and the European Commission a right of access to data to the extent necessary for monitoring and assessing compliance.

[27] See also Cauffman and Goanta 'A New Order: The Digital Services Act and Consumer Protection', 2021, 12 EJRR 758, 770 et seq.

V. Implementation, Cooperation, Penalties, and Enforcement (Chapter IV)

Chapter IV contains provisions on the competent authorities, including details about the role and powers of the national Digital Services Coordinators (cf. Section 1, Art. 49 et seq. DSA), rules on cooperation and enforcement (Sections 2 to 5, Art. 56 et seq. DSA), and provisions on delegated and implementing acts (Section 6, Art. 87 et seq. DSA). Enforcement of the DSA is generally up to the member states (cf. Art. 49 para 1 DSA, Recital 79). To this end, all member states must appoint an independent Digital Services Coordinator (see Art. 38 DSA) and provide him or her with sufficient powers and means to ensure effective investigation and enforcement (cf. Art. 51 DSA).[28] The Digital Services Coordinator can be addressed by the recipients as well as other organizations mandated to exercise their rights with complaints against providers of intermediary services alleging an infringement of the provisions stated in the DSA (cf. Art. 53 DSA).[29] Generally, the competent supervisory authority is determined by the location of the main establishment of the provider of intermediary services, with the exception of supervision and enforcement against providers of very large online platforms and of very large online search engines, which lies with the European Commission (cf. Art. 56 para 1 to 3 DSA). Comparable to the GDPR,[30] a European coordination committee (European Board for Digital Services) is established to promote cooperation between supervisory authorities (cf. Art. 61 to 63 DSA).

VI. Outlook

The adoption of the DSA represents an important milestone in the broader European legal framework for a modern and secure digital environment that ensures not only the safety of users online, but also the protection of fundamental rights, including the freedom of speech as well as fair and contestable markets. This was also highlighted in European Commission President Ursula von der Leyen's reaction following the political agreement on the DSA: "*Today's agreement on the Digital Services Act is historic, both in terms of speed and of substance. The DSA will upgrade the ground-rules for all online services in the EU. It will ensure that the online environment remains a safe space, safeguarding freedom of expression and opportunities for digital businesses. It gives practical*

[28] On the powers of the Digital Services Coordinator see also Cauffman and Goanta 'A New Order: The Digital Services Act and Consumer Protection', 2021, 12 EJRR 758, 772.

[29] See also DSA, recital 110.

[30] See Art. 70 GDPR.

effect to the principle that what is illegal offline, should be illegal online. The greater the size, the greater the responsibilities of online platforms. Today's agreement – complementing the political agreement on the Digital Markets Act last month – sends a strong signal: to all Europeans, to all EU businesses, and to our international counterparts."[31]

[31] See EC Press Release, IP/22/2545, 23 April 2022.

Potentials and Limits of Filter Technology for the Regulation of Hate Speech and Fake News

Martin Steinebach

I. Introduction

Today, social media is an integral part of opinion-forming for large parts of the population.[1] For many, the internet has replaced traditional media such as news broadcasts or newspapers as their primary source of information. At the same time, a risk arises here that did not exist with conventional media: social media is based on user-generated content, which means that in principle anyone can share any statement without being regulated.

This freedom is abused to spread disinformation and hate speech, i.e., to deliberately circulate false statements. Closely related to this is misinformation, in which the statements are incorrect, but it is not intentional and instead due to ignorance. Disinformation poses numerous risks, which can be of a political,[2] social,[3] or even health-related[4] nature, among others. Hate speech[5] is always an attack on the dignity of individuals or entire groups of people. Therefore, measures are sought to contain and combat such messages.[6]

Intuitive here is the idea of installing a filter that automatically recognizes such messages and stops them from spreading. This is closely related to approaches to

[1] Acknowledgment: This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE as well as within the German Federal Ministry of Education and Research project DYNAMO.

[2] E.g., Tucker et al., 'Social media, political polarization, and political disinformation: A review of the scientific literature', 2018.

[3] E.g., Chenzi, 'Fake news, social media and xenophobia in South Africa', 2021, 19 (4) African Identities 502.

[4] E.g., Melchior and Oliveira, 'Health-related fake news on social media platforms: A systematic literature review', 2022, 24 (6) New Media & Society 1500.

[5] E.g., Paz, Montero-Dıaz, and Moreno-Delgado, 'Hate speech: A systematized review', 2020, 10 (4) Sage Open 2158244020973022.

[6] Shu et al., 'Combating disinformation in a social media age', 2020, 10 (6) Wiley, Interdisciplinary Reviews: Data Mining and Knowledge Discovery e1385.

blocking image content of child abuse,[7] which is already used on numerous social media platforms. But upload filters,[8] which are primarily used to protect against copyright infringement, also have many similarities.

This paper is a written reproduction of a presentation given on the presented subject. Its structure follows the presentation and aims to also address the follow-up discussion. For a more technical approach, it is recommended to access the numerous survey papers about hate speech and disinformation detection as well as social media filtering, some of which are referenced in this paper.

We will first introduce a generic structure of an online filter in the section "Basic Filter Concepts". Then we briefly introduce some technical approaches to identify content to be filtered. The core of the presentation is the section about the limits of filtering technology. Here we discuss challenges caused by the amount and content of online messages.

II. Basic Filter Concepts

A filter for online content consists of several components. Abstractly, at least the following must be present:

- **Reference Database** Based on this reference data, the filter decides how to react. Usually there is a list, in which the cases are stored, to which the filter reacts with further measures. Depending on the strategy for detection, the data can be the base of a set of robust or cryptographic hashes, or it can be used as training data for machine learning. Similarly, the database can represent a list of signal words for hate speech detection.
- **Decision Procedure** At the core of the filter, the question of whether an incoming datum (i.e., an image, text, or video) has a property that requires a reaction that must always be resolved. This classification can be done in many ways. Its reliability depends on how high the degree of freedom of the incoming data is in relation to the reference database. If only content is to be reacted

[7] Lee et al., 'Detecting child sexual abuse material: A comprehensive survey', 2020, 34 Forensic Science International: Digital Investigation 301022.

[8] Steinebach, 'Ausprägungen von Uploadfiltern' in *Selbstbestimmung, Privatheit und Datenschutz*, 2022.

to, which exists exactly in such a way as stored in the database, a reliable decision is possible. If also content *similar* to the one stored in the databased is to be identified, the decision becomes more challenging.

- **Reaction Mechanism** Based on the decision of the classification procedure, the filter must be able to react to the content. This reaction can be manifold depending on the deployment scenario. Content distribution might be slowed down or even blocked in severe cases. A case of perceived hate speech on a social network is likely to be referred to a moderator, who will make a final decision on whether to block or distribute it.

III. Known vs. Unknown Content

There are different approaches that deal with the handling of content. When filtering social media content, the task is to decide about a given message under examination. This message can contain text, video, image, or sound as well as a combination of these elements. In this context, it is important to mention that the procedures differ significantly for different media types. A solution for images is not easily transferable to texts. At the top level, this can be divided into two groups:

- **Recognition** Here it is assumed that a content is already known, e.g., from a previous examination. It is to be re-identified and matched with some information about it stored in a database. This can be done, for example, using cryptographic (common for texts) or robust (common for media such as audio, video, and images) hashes. It must be emphasized that this task is about identifying the content itself and not about extracting further information from it. For example, the task of recognizing a depicted person in an image,[9] does not count towards the re-identification discussed here.
- **Classification** In this case, it is not assumed that the content is already known. Instead, metadata is automatically generated, e.g., by matching the image to reference images with similar features. In this way, filtering for rele-

[9] Gong et al., 'The re-identification challenge' in *Person re-identification*, 2014.

vant features can be performed. Numerous approaches exist on how to classify images.[10] For image classification, trained deep learning networks are known to give the best results.[11] There are several general networks available that can automatically tag or annotate images. The actual decision is then based on the obtained annotations. For text messages, there are also multiple solutions for disinformation[12] or hate speech[13] detection, all based on natural language processing (NLP) methods.

As already mentioned, decisions on the content will be more reliable for recognition tasks than for classification tasks. In the following, we briefly mention a limited number of mechanisms commonly used for content decisions.

1. Cryptographic Hash Functions

Cryptographic hash functions[14] are common primitives of security protocols, with many applications that have long been known in IT security.[15] They compute fixed-length hash values from arbitrary-length information. They must meet a number of requirements, including the following:

- Efficiency: They must be computable with little effort.
- Collision resistance: It must be extremely unlikely to find two pieces of information that have the same hash value.
- One-way function: It must be practically impossible to find the information associated with a hash value.

These properties mean that cryptographic hash functions are only suitable for recognizing identical copies of a piece of content, for example, a photo or a video. As soon as even minimal changes occur to the file storing the information, the hash is a

[10] See e.g., Kamavisdar, Saluja, and Agrawal, 'A survey on image classification approaches and techniques', 2013, 2 (1) International Journal of Advanced Research in Computer and Communication Engineering 1005.

[11] Druzhkov and Kustikova, 'A survey of deep learning methods and software tools for image classification and object detection', 2016, 26 (1) Pattern Recognition and Image Analysis 9.

[12] Bevendorff and others, 'Overview of PAN 2020: Authorship Verification, Celebrity Profiling, Profiling Fake News Spreaders on Twitter, and Style Change Detection', Avi Arampatzis and others, 2020.

[13] Rangel et al., 'Profiling Hate Speech Spreaders on Twitter Task at PAN 2021', Guglielmo Faggioli et al., 2021, in Bevendorff et al., 'Overview of PAN 2021: Authorship verification, profiling hate speech spreaders on Twitter, and style change detection', 2021.

[14] See e.g., Katz et al., *Handbook of applied cryptography*, CRC press, 1996.

[15] Damgård, 'Collision free hash functions and public key signature schemes', 1987.

completely different one. All that is needed is to save the file with a lossy compression algorithm such as JPEG for pictures or h.264 for videos. The quantization that takes place leads to changes in the file and a break in the hash.

Therefore, corresponding procedures are not suitable for enabling re-identification within online content filters. Due to their properties, they play a role in integrity checks or in the search for digital duplicates, for example. Since changes to the content must always be expected in the context of filters, whether deliberately to circumvent the filter or unconsciously through processing steps, so-called "robust" hash procedures are necessary that are resistant to slight changes. If texts need to be re-identified, cryptographic hashes are often applied. However, they are not used as a hash over the complete text; instead, individual text passages are hashed in the sense of a continuous window function. In this way, sections of text can also be recognized.[16]

2. Robust Hash Functions

Several robust or perceptual hashes are known for different media types, offering different degrees of robustness. Since there are too many algorithms to mention here, we recommend surveys such as the one by Haouzia et al.[17,18] or Neemila and Singh.[19] Methods also exist for audio[20] and video streams,[21] as well as for text data.[22]

Robust hash functions extract perceptually relevant features from multimedia content for identification purposes. They have to fulfill a number of requirements. The most important are noted here:

- Distinction: Perceptually different pieces of media data should have different hash values.
- Robustness: The robust hash values should have some perceptual invariance, i.e., two pieces of media data that are similar to an average viewer/listener in terms of perception should be similar.

[16] Steinebach et al., 'Robust hash algorithms for text', 2013.
[17] Haouzia and Noumeir, 'Methods for image authentication: A survey', 2008.
[18] (1) Multimedia tools and applications 1.
[19] Neelima and Singh, 'A short survey on perceptual hash function', 2014, 1 ADBU Journal of Engineering Technology.
[20] Jaap Haitsma, Ton Kalker, and Job Oostveen, 'Robust audio hashing for content identification', 2001, vol 4.
[21] Job C Oostveen, Ton Kalker, and Jaap Haitsma, 'Visual hashing of digital video: Applications and techniques', 2001, vol 4472.
[22] Steinebach et al. (fn. 16).

- Security: The features must survive attacks that directly target the feature extraction and subsequent processing steps. Similar to cryptographic hash functions, the robust hash values shall be evenly distributed over all possible media data and pairwise statistically independent for two media data that differ in perception.

Robustness carries the risk of information leakage: If two images are very similar, their hashes are also similar. The distinction only goes so far that two similar images do not have an identical hash, but both hashes are more similar than the hashes of two images with different content. As an example, portrait photos with a human face in the center and a light, monochrome background all have a similar robust hash structure. This leads to false positives for a robust hash function that is higher than expected given the theoretical number space spanned by a hash.

The reliability of robust hash methods in recognition is high. For example, our image hash ForBild[23] has a false positive rate of 0% and a false negative rate of 0.2% in the test conducted.

3. Feature Matching

Methods that implement feature matching are characterized by a higher resistance to image changes than that of robust hash methods. They can survive rotation and distortion well, and cropping the image is often unproblematic. The methods are based on detecting so-called key points at several locations in an image with a detector and extracting descriptors with a feature extractor. In a further step, the feature comparison, the found features are compared with features of another image. If both images now contain the same object, the features should ideally be measurably similar. A feature itself is defined as an "interesting" part of the image. What exactly is understood as an "interesting" part of the image varies depending on the feature detector. The part of the image where a feature is extracted is often either an isolated point, a continuous curve, or a connected area.

The Scale Invariant Feature Transform (SIFT)[24] algorithm is one of the best-known and most commonly used feature detectors. The Speeded Up Robust Features (SURF) detector[25] is partly inspired by SIFT and is an attempt to be faster and more robust than SIFT.

[23] Steinebach, Liu, and Yannikos, 'Forbild: Efficient robust image hashing', 2012, vol 8303.

[24] Lowe, 'Distinctive image features from scale-invariant keypoints', 2004, 60 (2), International Journal of Computer Vision 91.

[25] Bay, Tuytelaars, and Gool, 'Surf: Speeded up robust features', 2006.

The method is generally more complex than robust hash methods and is therefore only used when the application requires resistance to rotation and distortion. These methods can also be helpful when recognizing individual parts of an image or inserting one image into another. For example, we detected 99% of the inserted image objects and 100% of the image backgrounds in which the objects were inserted in the context of the detection of image montages.[26]

4. Natural Language Processing

This term covers numerous approaches to extract information from unstructured text. Methods range from statistical analysis of the occurrence of previously determined signal words to the recognition of context or authorship using machine learning. Which approaches are used here depends on the concrete application:

- If a filter is to prevent a once-blocked user from uploading content again using a new username, it is a task for authorship detection. The writing style of the blocked author is learned, and then new texts are checked for this style. For example, an accuracy of 79% was achieved by Halvani[27] for German-language texts. There are also methods to recognize already blocked content. As mentioned above, robust hash methods are available for text. Also, cryptographic hashes, usually applied by a sliding window, are used for recognition.
- If bots are to be prevented from spreading messages in a channel, a distinction must be made between bots and humans.[28] In the PAN Challenge, an international comparison of Natural Language Processing solutions, the average recognition rate for distinguishing between messages from a bot and a human in English was 86%.[29]

[26] Steinebach et al. (eds.), 'Desinformation aufdecken und bekämpfen – Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität', 2020; Steinebach, Gotkowski, and Liu, 'Fake News Detection by Image Montage Recognition', 2019.

[27] Halvani, Winter, and Pflug, 'Authorship verification for different languages, genres and topics', 2016, 16 Digital Investigation, p. 33.

[28] de Oliveira et al., 'Identifying fake news on social networks based on natural language processing: Trends and challenges', 2021, 12 (1) Information 38.

[29] Rangel and Rosso, 'Overview of the 7th author profiling task at PAN 2019: Bots and gender profiling in Twitter', 2019.

- If the task is to identify hate speech in general, approaches are needed that act more on content and are independent of the author. For example, the X-SONAR project was able to correctly detect content containing hate speech in Twitter messages in 85% of the cases.[30]
- In the corresponding PAN challenge,[31] the winner[32] succeeded in detecting disseminators of disinformation with an average accuracy of 77.8%. Solutions based on simple mechanisms such as n-grams and support vector machines[33] also achieved an accuracy of over 75%. The 2022 PAN challenge[34] provides multiple approaches for COVID-19 fake news detection.

The list of technical components with which a filter can process content can be extended at will. For example, classification can of course also be carried out for videos, and here too machine learning is now widely used.[35] Ultimately, any method that can automatically extract metadata from a piece of content can potentially be used in the context of a filter if the metadata is relevant to the filter's decision.

IV. Limits

So far in this work, we have introduced the concept of filters for social media content and discussed different technologies and approaches. In this section, we discuss a number of challenges and limitations which can occur when using automated filtering in real-world scenarios.

1. Handling of False Alarms

In the discussion about filters for social media content, their technical characteristics and properties are often not considered. Systems based on recognition have significantly lower error rates (usually less than one per cent) than those that classify content using machine learning. Here, the error rates are often over 10 per cent. If the two

[30] Vogel, Regev and Steinebach, 'Automatisierte Analyse Radikaler Inhalte im Internet', 2019, INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik– Informatik für Gesellschaft.

[31] Rangel et al., 'Overview of the 8th author profiling task at PAN 2020: Profiling fake news spreaders on Twitter', 2020.

[32] Citebuda2020ensemble.

[33] Vogel and Meghana, 'Fake News Spreader Detection on Twitter using Character N-Grams', 2020.

[34] Nakov et al., 'The CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection', 2022.

[35] Tran et al., 'Video classification with channel-separated convolutional networks', 2019.

approaches are not clearly separated from each other in the discussion, unrealistic expectations regarding the performance filters can arise. Today, the detection of "hate speech" or "fake news" cannot work with the same reliability and therefore automation as, for example, ContentID[36] in YouTube, which recognizes content.

At the same time, however, the requirements on the reliability of filters for "hate speech" and "fake news" are high when applied to real-world scenarios. A corresponding filter would have to examine and evaluate all messages in social networks. In 2022, there were more than 800 million tweets per day on Twitter.[37] In the Meta (Facebook, Whatsapp, Instagram, ...) group, there are 100 billion messages per day.[38] Via Telegram, 15 billion messages are sent per day.[39]

Thus, the error rate of these procedures becomes even more important, as the operators of social media are likely confronted with a large number of generated warnings from their filters, a large proportion of which are incorrectly classified.

If we assume, for example, that one out of every 1,000 messages contains "hate speech", then with roughly 90% correct detection, 900 correct alarms are raised for one million messages. But if 10% of the remaining 999,000 benign messages are misclassified, 99,000 false alarms would be raised. So for 900 correctly classified messages, there are 99,000 false alarms. If a system was fully automated, almost 10% of all messages would be incorrectly blocked. If employees double-check the messages, this would require a high use of human resources: given the numbers of the Meta group quoted above, almost 10 billion false alarms would have to be handled daily.

However, these values can be adjusted. Classification is usually decided with threshold values: an assignment is made when a trained network is certain about a given percentage. For example, it can be specified that a network must be 80% sure that a text is "hate speech" in order to classify it accordingly. This turns a percentage into a binary decision. If this threshold is changed, two things happen: the recognition rate of genuine hits drops because fewer hits exceed the threshold and lead to a report. At the same time, the number of examples falsely classified falls, as these also fail the threshold more often. The challenge is then to find a threshold that will detect fake news or hate speech with an acceptable reliability while producing minimal false alarms. Looking at the Meta example again: if we assume that 1 million daily false alarms are acceptable, the false alarm rate needs to drop to 0.001 percent. This is

[36] Solomon, 'Fair users or content abusers: The automatic flagging of noninfringing videos by content id on Youtube', 2015, 44 Hofstra L. Rev. 237.

[37] https://www.businessdit.com/number-of-tweets-per-day/.

[38] https://bloggingwizard.com/facebook-messenger-statistics.

[39] https://www.demandsage.com/telegram-statistics.

10,000 times lower than current research results. If we expect the true positive to linearly move in the other direction, we drop from 90% to 0.009%, basically disabling detection.

Of course, this is a simplified look at the problem, but it shows why fully automated or event semi-automated filtering is challenging today. It is more likely that filters are very selective, looking only at messages that are widely distributed, that are flagged as problematic by users, or that come from sources already known to be likely to spread hate speech or fake news due to previous detection.

2. Topic Dependence

When using filters and evaluating them, a critical question must always be asked: Are the results independent of the topic of the news? Does the filter recognize hate speech and fake news, or does it "only" recognize homophobia and misogyny, for example, but then fails to recognize antisemitism as well? Similarly, a filter based on machine learning that has been trained with disinformation about Covid-19 and is good at recognizing it may not be able to do the same for disinformation about the Ukraine war.

This is because often when learning a characteristic from training data, topic and style are mixed by the AI model. For example, one type of disinformation may use more technical language to misrepresent or incompletely reflect facts. For another topic, it may be more likely to appeal to personal fears that are more abstract and in the future. Training recognition that reliably detects disinformation regardless of the subject of the argument is currently only possible to a limited extent. As a consequence, it is very likely that disinformation campaigns on a new topic cannot be recognized well at first. Only when corresponding messages have been marked manually and can be used for training can automated recognition be successful.

It must also be taken into account that the language in which texts are available plays an important role when using NLP (Muhammad Arif et al., "CIC at CheckThat! 2022: Multi-class and cross-lingual fake news detection" [2022] Working Notes of CLEF). Common languages such as English, Spanish, German, and French are easier to handle than less common languages. This applies both if the texts are to be analyzed directly in the language at hand and if translation into another language is to take place beforehand. Accordingly, local disinformation that may only concern a smaller country with a less widely spoken language can be more difficult to detect because the basic technical image of the language is incomplete.

3. Unstructured Data

Disinformation and hate speech are usually spread in social media in the form of short texts, often supported by images or other media. For a computer-based analysis, such

data is unstructured.[40] The detector must first recognize where relevant statements occur in the data. If it fails in this, the actual analysis cannot be successful. If disinformation is to be recognized based on content analysis, the claimed information (true or false) must first be derived. To do this, it must be recognized which parts of a message can be checked at all. This is called check-worthiness detection.[41]

This can be especially challenging when information and disinformation are mixed in a message. Even a single sentence can include true and false information: "Mr. Doe, the president of our country, today provided evidence that all members of public transport services are reptiles" may start with a true statement about Mr. Doe being the president but continue with disinformation about the members of the public transport services. A detector must now recognize that (a) the part about the public transport members needs to be checked and (b) that this statement is sufficient to raise a disinformation alarm.

Images and other media types are also unstructured data with respect to filtering unwanted content. Filtering images showing harmful symbols is a challenge well-known in image processing,[42] which is relevant for hate speech detection. The presence of certain symbols can be sufficient to filter a message. But in other cases, an understanding of an image and the relation to the text next to the image can be necessary to identify hate speech or disinformation. A piece of text saying "A photo of our president" is unproblematic. A photo of a monkey is also okay. A combination of both may be an indicator for content to be blocked. The text "A photo of Mr. Smith after his vaccination" together with the photo of a dead patient in a hospital can well be part of a disinformation campaign.

To identify such combinations, the detector needs to understand what is shown in the pictures and relate the derived content to the message. Only then can it decide the nature of the message. To successfully do so, it also needs to be able to attribute the relevant information of, for example, a photo to a message. If the dead patient in the example above is accompanied by (living, healthy) relatives standing next to the bed, it needs to select the dead patient as the relevant part of the photo for its decision.

[40] Feldman, Sanger, et al., 'The text mining handbook: Advanced approaches in analyzing unstructured data', 2007.

[41] Hansen et al., 'Neural Weakly Supervised Fact Check-Worthiness Detection with Contrastive Sampling-Based Ranking Loss', 2019.

[42] Zheng, Liu, and Daoudi, 'Blocking objectionable images: Adult images and harmful symbols', 2004, vol 2.

These are just of few examples of existing challenges and limits of disinformation and hate speech detection. The list is neither complete nor very systematic. The goal is to show that the challenges are far from solved and that expectations of automated filtering should be conservative.

V. Discussion

The importance of social media for everyday communication and the associated reach as a dissemination channel for information makes it obvious that controlling content within the framework of what is legally prescribed is an important challenge. Users need to be protected from deliberate manipulation of their opinions by the spread of false information. Similarly, they need to be protected from becoming victims of hate and bullying online.

Due to the large amounts of news that are spread daily, technical support is indispensable. It will not be possible to meet this challenge purely manually, even if the companies behind the social media are willing to invest a lot of resources in it.

At the same time, however, it is necessary to take a sober look at what the real state of the technology is. Laboratory experiments on message detection are usually very different from daily practice. The topics of disinformation campaigns can change quickly. Language on social networks is dynamic and fast-moving. New terms and expressions are constantly emerging, and it is difficult to include them all in recognition models. Irony and sarcasm are widespread and still pose major challenges to recognition today. Facts can be misused out of context without being wrong in and of themselves.

This challenge leads to error rates that result in a disappointing performance of corresponding filters for users. Currently, such filters should be used in a hybrid fashion. Unknown content should only be evaluated in terms of its likelihood of being problematic. Very likely problematic cases should then be evaluated by humans. In this way, a pre-sorting can take place based on the machine's evaluation, which ideally quickly identifies the particularly relevant cases. Through the interaction of humans and machines, a quick reaction is then possible. Once the relevant content has been identified as needing to be filtered, the software can reliably recognize it if it occurs again. A hybrid solution thus directs attention to relevant new and potentially problematic cases, leaves the final decision to the human, and then implements this decision in case of recurrences without involving the human again.

Of course, such hybrid systems are not without risk. Relevant content will not always be detected during presorting. Humans can make mistakes in their decisions. Recognition will make fewer errors than the classification of unknown content, but

it is not error-free. And of course, there can be deliberate content obfuscation attacks whose goal is to spread problematic content without being detected.

Freedom of Speech goes Europe - EU Laws for Online Communication

Antje von Ungern-Sternberg[1]

I. Introduction

"Direct hit, freedom of the press sunk". This is how the *Frankfurter Allgemeine Zeitung* (FAZ) titled an article on the European Union's *Digital Services Act* in January 2022, fearing that the new piece of legislation would allow large digital corporations to censor unpopular press articles.[2] On closer inspection, this fear hardly seems justified; on the contrary, it forms a fine example of classic reservations about legislation in the European Union. One could reformulate the unease as follows: Once again, a classic domain of national law, i.e. the democratic formation of opinion, is being Europeanised. And indeed, the European Union is developing extensive legislative activities on issues of digitalisation and the digital economy with great self-confidence. In 2021, the European Commission proclaimed "Europe's Digital Decade".[3] The Digital Services Act (DSA), already passed to deal with harmful content and practices on the internet, is only one particularly prominent building block. Other legislative projects concern

[1] This contribution is based on a (German) talk given at the XXI. Walter Hallstein-Kolloqium on 3 March 2022, to be published in Kirchner, Heger, Hofmann, and Kadelbach (eds.), Die Digitalisierung im Recht der EU, Baden-Baden 2023, forthcoming.

[2] Hanfeld, 'Volltreffer, Pressefreiheit versenkt', FAZ.NET, 22 January 2022, https://www.faz.net/aktuell/politik/staat-und-recht/das-neue-digitalgesetz-der-eu-stellt-die-pressefreiheit-in-frage-17744847.html. The lurid presentation is in stark contrast to the only brief suggestion of the legal basis on which press freedom could be disregarded: It is about the possibility of digital platforms to set up general terms and conditions and the obligation to 'respect' media freedom in doing so. See Art. 14 (1) and (4) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/32/EC (Digital Services Act).

[3] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_de.

competition in the digital sector,[4] artificial intelligence,[5] or the use of data.[6] Although the projects, which are called "laws" and no longer just "regulations", focus primarily on economic actors such as digital platforms, they also affect the democratic formation of opinion, the normative framework of which has so far been determined primarily by the member states. In January 2022, the Commission also presented a draft for a "European Declaration on Digital Rights and Principles for the Digital Decade",[7] to be proclaimed by the European Parliament, the Council, and the Commission. This also addresses participation in the digital public sphere, advocates digital technology to promote civic and digital participation, and commits to freedom of expression and information in the online environment, as well as combating harmful illegal content and disinformation.[8]

Is democratic opinion-forming now in good hands with the European Union? This could be doubted, for example, by those who accuse the European Union of institutional deficits in the protection of fundamental rights[9] or criticise the content of European solutions because they take too much account of the interests of the digital economy at the expense of the press[10] or restrict freedom of expression and information too much in favour of data protection.[11] This contribution is less sceptical. It elaborates – after a look at characteristics of opinion formation in the digital space (II.) – how the law of the European Union actually increasingly covers opinion formation in the digital space and thereby responds to central challenges for online discourse (III.). This is followed by an outline of the principles that the ECJ has so far developed with regard to

[4] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[5] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Acts, 21 April 2021, COM/2021/206 final.

[6] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) with regard to data held by public sector bodies (abbreviation according to the English acronym also DGA); Proposal for a Regulation of the European Parliament and of the Council on harmonising rules for fair access to and use of data (Data Act) of 23 February 2022, COM/2022/68 final, with regard to business data.

[7] Draft of 26 January 2022, COM/2022/28 final.

[8] Chapter IV of the draft of 26 January 2022, COM/2022/28 final.

[9] On the history of reservations in the case law of the Federal Constitutional Court Sauer, Staatsrecht III, 7th ed. 2022, § 9, marginal no. 26 et seq. Furthermore, a newspaper article by the then judge of the Federal Constitutional Court, Johannes Masing, caused a stir, in which he warned of institutionally inadequate protection of fundamental rights by the ECJ with regard to the planned General Data Protection Regulation, with detrimental effects for data protection, freedom of opinion or the general right of personality: 'Ein Abschied von den Grundrechten' SZ, 9 January 2012, p. 10.

[10] In this sense for example Hanfeld (fn. 2).

[11] On this criticism in response to Case C-131/12 *Google Spain,* 2014, for example Kulk and Zuiderveen Borgesius, 'Google Spain v. González: Did the Court Forget about Freedom of Expression?', EJRR 2014, 389.

freedom of opinion as a central fundamental right of democratic opinion-forming (IV). These are similar to the ideas of Basic Law. There is therefore much to suggest that the ECJ will continue to sensibly develop this case law in cooperation with national courts in the future.

II. Opinion Formation in the Digital Space

First, it should be explained which characteristics of the exchange of opinions in the digital space raise the question of legal reactions. Digitalisation has led to a shift of opinion formation in the online world. Classic analogue media and forums (press, radio, marketplace, regulars' table) are increasingly being supplemented or replaced by their counterparts on the internet. There, alongside the digital offerings of the press and broadcasting, new formats such as blogs, podcasts, or influencer videos can be found, as well as new information platforms such as *WhatsApp*, *YouTube*, *Facebook, Instagram, TikTok,* or *Twitter*,[12] where people not only communicate with each other privately or in business, but also politically. The countless possible pieces of information are conveyed to the respective user by algorithms – for example, by a search engines like *Google* or by algorithms of the internet platforms that display and sort news or messages according to (presumed) individual user preferences. This digitalisation of communication holds huge opportunities, especially because citizens can access and exchange relevant information easily, quickly, reliably, and across borders. Nevertheless, in the digital space, there is also a particular danger that harmful content will be disseminated (1.) and that problematic techniques will be used (2.). Furthermore, the question arises as to what role the powerful information platforms have and should have in the formation of opinion in the digital space (3.).

1. Harmful Content

Harmful communication content such as hate speech or disinformation can be found in both the analogue and digital worlds. But they pose particular problems in the digital space.[13] Firstly, the victims are often more affected because the content spreads easily, quickly, and widely by digital means; because the authors, who do not always act under

[12] S. approx. https://de.statista.com/statistik/daten/studie/800623/umfrage/nutzung-von-sozialen-medien-nach-plattform-in-deutschland/.

[13] See for instance Fichter (ed.), Smartphone-Demokratie. #Fake News, #Facebook, #Bots, #Populism, #Weibo, #Civic Tech, 2016; Kaspar, Gräßer, and Riffi (eds.), Online Hate Speech, 2017; Bayer et al., Disinformation and propaganda - Impact on the functioning of the rule of law in the EU and its Member States, 2017; Hohlfeld et al. (eds.), Fake News und Desinformation, 2020; Möller, Hameleers, and Ferreau, Typen von Desinformation und Misinformation, 2020, www.die-medienanstalten.de.

clear names, may be difficult to find and hold responsible (especially abroad); and because even in the case of clearly illegal content, active action by the respective information provider is still required to block or remove harmful content. Secondly, the logic of action of the relevant communication participants has changed. While the press and broadcasting are obliged to observe journalistic standards of care, this does not generally apply to other communication participants. On the internet, it is no longer only the traditional media that provide content, but practically all users – from extensive texts, pictures, or videos to small statements in the form of *WhatsApp* messages, posts, likes, or retweets. The motives for this can be private, professional, commercial, political, or mixed. Think, for example, of the scientist who uses *Twitter* to draw attention to new research results, support political demands, cheer along with her football club, and promote her textbook. Standards, for example on the truth, objectivity, or impartiality of statements, do not exist, even if these can develop in specific areas – for example for commercial and political bloggers or influencers.[14] Furthermore, there are many indications that online communication can have a disinhibiting and radicalising effect on participants under certain circumstances.[15] Thirdly, internet platforms also play a key role. They are predominantly private, profit-oriented companies that finance themselves through advertising or contributions and are oriented towards the preferences of their users and not towards Habermas' ideal of discourse. Since many users are particularly interested in lurid and polarising content, the mediation algorithms often accommodate these preferences.[16] The fact that money can be made precisely with harmful content was recently confirmed by the revelations of whistleblower Frances Haugen about Facebook.[17]

The consequences of harmful online content can be severe. Hate speech – a term that encompasses both unlawful and lawful but harmful statements of disparagement

[14] In Germany, this is currently primarily the case for the economic sphere, see Peifer, 'Die neuen Transparenzregeln im UWG (Bewertungen, Rankings und Influencer)', GRUR 2021, 1453.

[15] Strobel, 'Die Grenzen des Dialogs' in: Kaspar, Gräßer, Riffi (eds.) (fn. 13), 29, 32 ; Schmitt, 'Online Hate Speech: Definition und Verbreitungsmotivationen aus psychologischer Perspektive' in ibid., 51, 52; Kaspar, 'Hassreden im Internet - Ein besonderes Phänomen computervermittelter Kommunikation?' in ibid., 63.

[16] S. Kühling, '"Fake News" und "Hate Speech"- Die Verantwortung der Medienintermediäre zwischen neuen NetzDG, MStV und Digital Services Act', ZUM 2021, 461, 463 .

[17] Concrete examples relate to the spread of violence and the promotion of criminal offences (trafficking in human beings and organs) or health hazards (eating disorders, vaccination scepticism), https://netzpolitik.org/2021/facebook-leaks-whistleblowerin-erhebt-schwere-vorwuerfe-gegen-facebook/. Problematic content is also attractive on the other platforms, be it the Spotify podcast by Joe Rogan, who spreads disinformation about Covid 19 or racist vocabulary in front of an audience of millions https://www.spiegel.de/kultur/joe-rogan-entschuldigt-sich-fuer-rassistische-beleidigung-in-seinem-podcast-a-1bf7e0e8-daba-425f-9521-3304c52d7496 or Amazon, which recommends all kinds of books by anti-vaccinationists on the subject of 'Corona' or 'vaccination', https://netzpolitik.org/2021/desinformation-im-netz-die-virenschleuder-amazon/.

and denigration, such as insults, rape fantasies, death threats, or incitement against certain groups of the population (typically women, minorities, politicians) – is problematic not only because it can constitute criminal offences or prevent a factual debate, but also because it can drive the victims completely off the net or out of the public eye and deter them from taking up political office.[18] Disinformation can also have serious consequences – in politics as well as in private life. Those who believe that the 2020 election was "stolen" from the US Republicans may join the armed storming of the Capitol; those who learn about alternative treatments (e.g., bleach) against the coronavirus may try them out. Incidentally, the use of such fake news is often not aimed at concrete behaviour, but rather at fuelling social conflicts or weakening trust in democratic institutions.[19]

2. Problematic Communication Techniques: Exploring, Deceiving, Manipulating

Problematic content is joined by problematic communication techniques. On the one hand, communication participants can be researched more and more intensively due to the intelligent evaluation of large amounts of data, thus being provided with specific content in a targeted manner and, if necessary, manipulated. Digital platforms, for example, can make relatively precise statements about the characteristics and preferences of their users due to the diverse interactions. This knowledge can then be used for precisely tailored – commercial or political – advertising to finely defined groups of recipients, i.e., for microtargeting.[20] Now, customised advertising is not illegitimate per se. What is problematic is the unauthorised processing of data and, in particular, the asymmetry of information and the manipulative potential that unilateral knowledge of the

[18] The Federal Constitutional Court has also recently identified this danger: 'Especially under the conditions of the dissemination of information through 'social networks' on the internet, an effective protection of the personal rights of public officials and politicians is in the public interest, above and beyond the significance for the persons concerned, which can increase the weight of these rights in the balancing process. For a willingness to participate in the state and society can only be expected if those who engage themselves and contribute publicly are guaranteed sufficient protection of their personal rights', BVerfG, NJW 2022, 680, 683 marginal no. 35 *Renate Künast.*

[19] See for example DiResta et al., 'The Tactics & Tropes of the Internet Research Agency', U.S. Senate Documents 10-2019, https://digitalcommons.unl.edu/senatedocs/2/; Howard et al., 'The IRA, Social Media and Political Polarization in the United States', 2012-2018, U.S. Senate Documents 10-2019, https://digitalcommons.unl.edu/senatedocs/1/.

[20] Fichter, 'Big Data im Wahlkampf - Mythos oder Waffe?' in Fichter (ed.) (fn. 13), 96; Blasi Casagran and Vermeulen, 'Reflections on the murky legal practices of political micro-targeting from a GDPR perspective', 2021, 11 International Data Privacy Law 348.

internet user by platforms (or data traders) entails.[21] This is especially true if the advertising specifically seeks to exploit particular weaknesses of the addressees, such as the receptivity to certain messages in a particular emotional or psychological state. Moreover, public control of microtargeting is also difficult as long as it takes place in secret and parties cannot be politically taken to task for contradictory promises to different voter segments, for example.

On the other hand, intelligent deception techniques are also increasingly raising questions in the digital space. With their help, users can deceive others about the content, authorship, or popularity of information. For example, image and sound documents can use artificial intelligence to fake actions and statements of real people ("deepfakes")[22] and be used for political purposes, such as a film showing the supposed surrender of Ukrainian President Selensky to Russian troops.[23] With "social bots" or other forms of coordinated, non-authentic behaviour, computer programmes imitate human behaviour on the net and thus not only simulate a human author of a statement, but also give it special credibility and reach.[24] For example, someone who poses online as a supporter of the *Black Lives Matter* movement would find a better hearing among like-minded people than a representative of the Kremlin.[25] It is true that there were many deception techniques in analogue times as well. For example, the term "astroturfing" – pretending to be a grassroots movement when it is only "artificial turf" – dates back to the 1980s.[26] But the digital architecture of the internet, where a person can create automated user accounts and posts and buy follower numbers and clicks in exchange for money, facilitates orchestrated deceptions and amplification of a message – even to the point of suppressing dissenting views.[27] All in all, the digital techniques of exploration

[21] See Zuiderveen Borgesius et al., 'Online Political Microtargeting: Promises and Threats for Democracy', 2018, 14 Utrecht Law Review 82, 87 paras; Towfigh and Luckey, 'Zielgruppenbasierte Ansprache von Wahlberechtigten durch politische Parteien', RW 2022, 61, 79 et seq.

[22] For example, Thiel, '"Deepfakes" - Sehen heißt glauben?', ZRP 2021, 202, 202 et seq. ; Kumkar and Rapp, 'Deepfakes - Eine Herausforderung für die Rechtsordnung', ZfDR 2022, 199.

[23] https://www.spiegel.de/netzwelt/web/meta-loescht-fake-video-das-wolodymyr-selenskyj-falsche-worte-in-den-mund-legt-a-5600045c-8057-4359-bd31-ee02c6e585d5.

[24] Von Ungern-Sternberg, 'Mehr Lauterkeit für Online-Kommunikation - Kennzeichnungspflicht zum Schutz der Demokratie vor Social Bots', RW 2022, 94, 99 et seq.

[25] To corresponding experiments https://www.wired.com/story/russian-black-activist-facebook-accounts/. On the detection of social bots and non-authentic behaviour in general, see Woolley and Howard, Computational Propaganda. Political Parties, Politicians, and Political Manipulation on Social Media, 2018; Howard, Lie Machines. How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives, 2020, 29.

[26] Scott, 'Ripping Up the Astroturf: Regulation Deceptive Corporate Advertising Methods', Iowa Law Review 105 (2019) 431, 436 .

[27] Von Ungern-Sternberg (fn. 24), 99.

and deception allow for new forms of targeted, manipulative influence that can be used not only for commercial but also for political purposes.

3. Key Position of the Platforms

Internet platforms play a key role in the dissemination of information on the net. Internet platforms in the broad sense are "internet-based applications that bring users of the internet into contact with each other and enable them to exchange goods, services and information",[28] and it is precisely the exchange of information (also as goods or services) that is of interest for the formation of opinion. The term platform can be used differently depending on the regulatory context.[29] In the broad sense used here, it includes particular social networks such as *YouTube*, *Facebook* or *Twitter,* i.e., platforms on which users share any content with other users or make it accessible to the public,[30] and furthermore messenger services such as *WhatsApp* or *Telegram,* which can serve individual and public communication. Also relevant for opinion forming are search engines such as *Google*[31] and trading platforms such as *Amazon,* which offer information as goods or services, and portals such as *Spotify*.

The special significance of these platforms now lies in the fact that they determine the rules of information dissemination themselves and thus exert a decisive influence on the formation of users' opinions, which they do in several respects. First, an effective fight against harmful content and problematic techniques is usually only possible with the help of the platforms. For example, it is particularly important for those affected by sensitive media coverage (e.g., on criminal offences, insolvencies, criticised employer

[28] Wagner, 'Die Haftung von Plattformen für Rechtsverletzungen', GRUR 2020, 329.

[29] For example, in addition to broadcasting, media law distinguishes telemedia according to § 2 para 1 sentence 3 of the Medienstaatsvertrag (MStV) (in particular non-linear online offers of goods and services, Martini in Gersdorf, Paal (eds.), BeckOK Informations- und Medienrecht, 38th ed., 1 February 2021, § 2, marginal no. 25), broadcasting-like telemedia according to § 2 para 2 no. 13 MStV (for example, non-linear online offerings of portals such as Netflix and Spotify), media platforms according to § 2 para 2 no. 14 MStV (which aggregate broadcasting, broadcast-like telemedia or the online press within the meaning of section 19 (1) into an overall offering) as well as media intermediaries pursuant to section 2 (2) no. 16 MStV (as a catch-all without the characteristic of an overall offer; this includes search engines, social networks, app portals for user-generated content, blogging portals and news aggregators, Martini in Gersdorf, Paal (eds.), BeckOK Informations- und Medienrecht, 38th ed., 1 February 2021, § 2, marginal no. 120). These would all be considered platforms in the broader sense.

[30] This is the legal definition in § 1 para 1 sentence 1 NetzDG (Network Enforcement Act).

[31] Online search engines (Art. 3 lit. j DSA) are distinguished from online platforms in the DSA. The latter are legally defined as a 'hosting service which stores and publicly disseminates information on behalf of a user, provided that this activity is not merely an insignificant and purely ancillary function of another service or an insignificant function of the main service [...]'. (Art. 3 lit. i DSA). Both fall – as do other platforms in the broader sense – under the general concept of an intermediary service (Art. 3 lit. g DSA).

practices)[32] that corresponding hits no longer appear in a *Google* or *YouTube* search, while the more elaborate and targeted search in specific online databases of newspapers or broadcasters, some of which are subject to a fee, is only undertaken by comparatively few interested parties and therefore also damages reputation or hinders resocialisation to a lesser extent. In addition, platforms can also detect harmful content and problematic techniques particularly well through their direct access and exposure, for example, *social bots* and other forms of coordinated inauthentic behaviour due to the circumstances of communication (e.g., communication networks, location, timing of contributions, monothematic content).[33] Furthermore, information platforms are, after all, not only bound by the legal guidelines on illegal or legal content and techniques. Particularly influential are the company's own privately set standards on permitted and prohibited types of communication on their platforms. For example, the major social networks such as *YouTube*, *Facebook* or *Twitter* also exclude (to varying degrees) hate speech, pornographic content, deepfakes, social bots, and election-related disinformation in their community standards.[34]

Beyond the question of harmful content and deceptive techniques, the platforms also determine which other rules apply to communication (e.g., obligation to use a clear name, human or machine moderation, disclosure of commercial or financial connections) and, in particular, which information is conveyed with which priority. The platforms display content to their users using search and sorting algorithms, such as *Google* search hits, the *Facebook* news feed, the *Twitter* timeline or *YouTube* recommendations. These algorithms can be primarily oriented towards user preferences or keep the degree of personalisation low, they can specifically limit or strengthen the reach of (supposedly) offensive content,[35] and they can advocate or oppose certain content in a special way as part of corporate policy. The extent to which this affects the formation of political opinion is controversial[36] and difficult to research – especially without access to the

[32] See for example the facts in Case C-131/12 *Google Spain,* 2014; BVerfGE 152, 152, *Recht auf Vergessen I*; BVerfGE 152, 216, *Recht auf Vergessen II.*

[33] Von Ungern-Sternberg (fn. 24), 96 et seq.

[34] The current status of the various Community Guidelines is available online, for example for Facebook, ('Community Standards' https://transparency.fb.com/de-de/policies/community-standards/; Twitter ('Twitter Rules' https://help.twitter.com/de/rules-and-policies/twitter-rules); YouTube ('Community Guidelines' https://www.youtube.com/intl/ALL_de/howyoutubeworks/policies/community-guidelines/).

[35] S. for instance Köver and Reuter, 'TikToks cap on disability', 2 December 2019, https://netzpolitik.org/2019/tiktoks-obergrenze-fuer-behinderungen/; Kayser-Bril, 'Automated moderation tool from Google rates People of Color and gays as "toxic"', 19 May 2020, https://algorithmwatch.org/en/automated-moderation-perspective-bias/.

[36] On the debate about filter bubbles and echo chambers, see von Ungern-Sternberg, 'Demokratische Meinungsbildung und künstliche Intelligenz' in: Unger, von Ungern-Sternberg (eds.), 'Demokratie und künstliche Intelligenz', 2019, 3, 10 et seq.; on the other hand, for instance Krafft et al., 'Wer sieht was?

companies' data. However, some studies support the intuitive assumption that the information offered also has an effect on opinion formation. For example, the *Facebook* button on one's own voting behaviour ("I voted", "I'm going to vote") increases the voter turnout of the addressees,[37] and a reduction of good or particularly bad news leads to a corresponding mood and a negative or positive communication behaviour of the addressees.[38] It therefore stands to reason that the EU legislature has taken on these powerful platforms, which hold a key position for the dissemination of information.

III. European Legislation on Opinion Forming in the Digital Space

For European legislation, it can be observed that at first it only covered the digital space of opinion-forming selectively and reflexively, but it has since become increasingly targeted. This is shown below – without claiming to be exhaustive – for the important regime of data protection (1.), for different approaches in the fight against harmful content and problematic communication technologies (2.), and for the new Digital Services Act (3.).

1. Data Protection as a Boundary and Basis for Democratic Opinion-forming

a) Data Protection as a Limit for Free Speech

Even if it is sometimes only apparent at a second glance, data protection affects democratic opinion-forming in many ways. Data protection has been standardised at the European level since 1995 with the Data Protection Directive (based on internal market

Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine', 2018, https://www.blm.de/files/pdf2/bericht-datenspende---wer-sieht-was-auf-google.pdf.

[37] On this Facebook experiment (conducted without users' knowledge): Bond et al., 'A 61-million-person experiment in social influence and political mobilization', 2012, 489 Nature 295; Grassegger, 'Facebook says its 'voter button' is good for turnout. But should the tech giant be nudging us at all?', 15 April 2018, https://www.theguardian.com/technology/2018/apr/15/facebook-says-it-voter-button-is-good-for-turn-but-should-the-tech-giant-be-nudging-us-at-all; Fichter, 'Die Schweiz wappnet sich für den Angriff aus dem Silicon Valley', 16 May 2018, https://www.republik.ch/2018/05/16/die-schweiz-wappnet-sich-fuer-den-angriff-aus-dem-silicon-valley.

[38] On this Facebook experiment (conducted without users' knowledge) Kramer et al., 'Experimental evidence of massive-scale emotional contagion through social networks', 2014, 111 Proceedings of the National Academy of Sciences 8788; Meyer, 'Everything You Need to Know About Facebook's Controversial Emotion Experiment', 30 June 2014, https://www.wired.com/2014/06/everything-you-need-to-know-about-facebooks-manipulative-experiment/.

competence)[39] and since 2016 with the General Data Protection Regulation (GDPR, additionally based on the specific competence for data protection and free movement of data of Art. 16 (2) TFEU).[40] In polemical exaggeration, data protection could be described as a natural adversary, or – less luridly – as the limit of open, democratic opinion-forming. This is because many forms of communication via the internet are in principle subject to data protection, insofar as they relate to personal data.[41] According to ECJ case law, the area exception for exclusively personal or family activities only extends to "private or family life", but not to publications that are "accessible to an unlimited number of persons", so that, for example, representations on the website of a church congregation about the characteristics and leisure activities of its employees are subject to data protection.[42] According to the case law of the ECJ, statements that can be perceived on the internet in general or by large groups (which can no longer be characterised as private or familial) must meet the requirements of data protection.[43] They, therefore, require sufficient legal ground in the form of consent or legal authorisation to process data.[44] As the ECJ already stated in *Google Spain* in 2014, this applies, in particular, to search engines when hits containing personal data are displayed.[45] It is, therefore, conclusively measured according to the GDPR when a delisting claim – a right to be forgotten – exists and when on the basis of Art. 6 (1) sentence 1 lit f GDPR, the legitimate interest of the search engine operator or the public in the display of a hit prevails.[46] In addition to search engines, according to the case law of the German Federal Supreme Court, rating portals in particular are also a case for data protection.[47]

---

[39] Directive 95/46/EC (Data Protection Directive).

[40] Regulation (EU) 2016/679 (General Data Protection Regulation, GDPR).

[41] For the elements of automated (or systematised) processing of personal data and the few exceptions, see Art. 2 (1) and (2), Art. 4 (1), (2) and (6) GDPR (or previously Art. 2 (a), (b) and (c), Art. 3 Data Protection Directive). Further provisions on data protection can also be found in Council Regulation (EC) 45/2001 (Data Protection by the EU), and Directive (EU) 2016/680 (Data Protection for the Prevention and Prosecution of Crime by the Police and the Judiciary).

[42] Area exception now under Art. 2(2) (d) GDPR; Case C-101/01 *Lindqvist,* 2003, para 47 (representations on the websites of a church congregation); confirmed in Case C-73/07 *Markkinapörssi,* 2008, paras 43 et seq. (publication of tax data in media); Case C-345/17 *Buivids,* 2019, para 43.

[43] Critical and in favour of a partial revision of the *Lindqvist case law*, according to which the private use of the internet and social media (and not the size of the circle of addressees) should be the decisive factor, Bäcker, in: Wolff, Brink (eds.), BeckOK Datenschutzrecht, 42nd ed., 1 November 2021, Art. 2 DSGVO marginal no. 21.

[44] Art. 6 para 1 GDPR (or previously Art. 7 Data Protection Directive).

[45] Case C-131/12 *Google Spain,* 2014, paras 21 et seq.

[46] The ECJ has already fleshed out the first delisting decision in further decisions, cf. Case C-136/17 *GC and Others v CNIL,* 2019 (delisting of sensitive data) ; Case C-507/17 *Google v CNIL,* 2019 (Spatial scope of delisting); Case C-460/20 *TU, RE v Google,* 2022 (Déréférencement d'un contenu prétendument inexact).

[47] BGH, MMR 2022, 202 *Jameda*; different view: Michel, 'Bewertungsportale und das Medienprivileg - Neue Impulse durch Art. 85 DSGVO?', ZUM 2018, 836.

The tension between freedom of expression and data protection is, however, somewhat more complicated in the GDPR. The special significance of freedom of expression is recognised in Art. 85 GDPR, which obliges the Member States to reconcile data protection on the one hand and freedom of expression and information on the other (paragraph 1) and to adopt national regulations for data processing for journalistic, scientific, artistic, or literary purposes (paragraph 2). The scope of this provision is partly unclear, but it decisively determines whether expressions of opinion relevant to data protection are to be assessed under the GDPR or under Member State law. While it is recognised that the member states can issue their own regulations on privileged data processing in journalism, science, art, and literature,[48] it is disputed whether there is also an additional member state competence to regulate general expressions of opinion relevant to data protection.[49] Furthermore, the question arises, for example, as to how the possible privileged circumstances are to be interpreted, i.e., whether one understands "journalistic" in the sense of established journalism (broadcasting, press) or if it also includes new, modern forms of information dissemination (bloggers, influencers). There is a case for a dynamic understanding, so that the Member States can take into account relevant modern forms of journalism (science, art, and literature)[50] and are not restricted in their primary competence of balancing freedom of science, art, and speech, with data protection as intended by Art. 85 GDPR. The ECJ also refers to "journalistic" broadly as those activities "which have as their purpose the dissemination to the public of information, opinions or ideas, by whatever means of transmission".[51] In concrete terms, it could, therefore, also constitute an activity for journalistic purposes if a citizen films – in his opinion unlawful – actions of the police and distributes the film on *YouTube* in order to draw society's attention to it.[52] This example illustrates that the ECJ will continue to be concerned with determining the scope of the privilege, e.g., distinguishing journalistic statements from institutional public relations or political advertising.[53] In addition, the ECJ will generally have the final say on the question of

[48] Art. 85 (2) GDPR. It is unclear here, for example, how a lack of notification of the standards to the Commission pursuant to Art. 85 (3) GDPR has an effect (probably irrelevant for the validity of the standard, see for example Pötters, in: Gola, Heckmann, DSGVO, BDSG, 3rd ed. 2022, Art. 85 GDPR, para 18) and how explicitly national standards must refer to the opening clause or data protection (see, for example, Lauber-Rönsberg, in Wolff, Brink (eds.), BeckOK Datenschutzrecht, 42nd ed. 1 November 2022, Art. 85, DSGVO, para 45 et seq., 48).

[49] Art. 85 (1) GDPR. This is not obvious from a systematic view of the norm; however, Nettesheim, 'Datenschutz und Meinungsäußerungsfreiheit', AfP 2019, 473, 474 et seq.; Lauber-Rönsberg, 'Zum Verhältnis von Datenschutzrecht und zivilrechtlichem Äußerungsrecht', AfP 2019, 373, 377; evidence for both sides at Lauber-Rönsberg (fn. 48), Art. 85 DSGVO, para 8.

[50] More generally, see Walter, § 13 Kommunikationsfreiheiten, in Grabenwarter (ed.) Enzyklopädie Europarecht - Europäischer Grundrechteschutz, 2nd ed. 2022, para 23.

[51] Case C-345/17 *Buivids,* 2019, para 53; see previously Case C-73/07 *Markkinapörssi,* 2008, para 56.

[52] Case C-345/17 *Buivids,* 2019, paras 55 et seq.

[53] Lauber-Rönsberg (fn. 48), Art. 85 GDPR , para 11.

whether national regulations violate European fundamental rights to data protection or freedom of opinion and information,[54] for example, the handling of a delisting claim against online archives of newspapers or broadcasters.[55]

b) Data Protection as the Basis for Free Speech

However, it would be too one-sided to see data protection only as a limitation of the free democratic formation of opinion. For it also strengthens the legal positions of communication participants in several respects, as will be explained by means of three examples. Firstly, a right of internet users to appear under a pseudonym or anonymously when communicating on the internet, and in any case not to have to comply with an obligation to use a clear name according to Facebook's terms and conditions, can be derived from data protection in connection with freedom of opinion.[56] This enables authentic expressions of opinion without fear of sanctions (i.e., also by persons who fear hostility on the net) and regardless of loyalty obligations (i.e., also by persons who report on internal practices or grievances in companies or authorities).

Secondly, data protection counteracts the exploration of the individual and thus knowledge asymmetries that can be exploited by the superior side. If a company knows a lot about a potential customer, it can use this knowledge – for example, about habits and preferences, mental and emotional states or distress – to submit tailored advertising and prices. The same is possible in non-commercial communication, for example when parties want to win voters, political officeholders want to improve their public image, or anti-democratic actors want to undermine trust in democratic institutions. Here, data protection sets limits to the screening of internet users and, in particular, erects high hurdles for the processing of sensitive data according to Article 9 of the GDPR, i.e., data on political opinions, religious beliefs, health, and sexual life. This ultimately protects the free formation of the opinion of the data subject concerned, as a secret communication advantage of the other side – superior knowledge on relevant topics,

[54] See, for example, Case C-101/01 *Lindqvist,* 2003, paras 84 et seq.; Case C-73/07 *Markkinapörssi,* 2008, paras 55 et seq.; Case C-345/17 *Buivids,* 2019, paras 50 et seq.

[55] Recognition of Member State decision-making competence only within the framework of Union law BVerfGE 152, 152, 170 et seq., *Recht auf Vergessen I*.

[56] See for example WD 10 - 3000 - 003/20; argumentation on old law, but with transferable arguments, Caspar, 'Klarnamenpflicht versus Recht auf pseudonyme Nutzung', ZRP 2015, 233; BGH, MMR 2022, 375, 379 et seq.: Prohibition of the pseudonym in the terms of use is unlawful after weighing the opposing fundamental rights positions of network operator and network user. It is disputed whether the same applies to the legal situation under the GDPR, which does not guarantee an explicit right to anonymity or pseudonymity and takes precedence over national law (current ban on the obligation to use a clear name under section 19 (2) of the Telecommunications Telemedia Data Protection Act). Hoeren, 'Klarnamenpflicht bei Facebook', MMR 2022, 375, 382 et seq., rightly advocates this right to pseudonymity.

promising argumentation, and advertising strategies as well as possible (cognitive, psychological, emotional) weaknesses of the addressees – can be prevented, reduced or at least made more transparent.[57]

Thirdly, data protection also grants the data subject a right to information about the type and scope of data processing from the data controller (Art. 15 GDPR). This claim can also be used to control public authority – similar to the information claims based on the Freedom of Information Acts of the Federation and the Länder.[58] The broad applicability of this right also makes it possible to inspect the activities of state examination offices[59] or parliamentary petition committees.[60] This control of public authority also enables the democratic formation of opinion so that data protection as a whole also forms a basis for the formation and expression of opinion.

2. Specific Action Against Harmful Content and Problematic Communication Techniques

The European Union is also attempting, within the framework of its competences, to take targeted action against specific harmful content and problematic communication techniques. Some regulatory approaches are presented below. It should also be noted that since 2015, the European External Action Service has also been carrying out very practical educational work against disinformation, especially of Russian and Chinese provenance.[61] In its regulation, the EU relies on instruments of voluntary self-commitment, i.e., soft law, as well as on classical laws.

---

[57] For an assessment of political microtargeting against the GDPR, see: European Data Protection Committee, Guidelines 8/2020 on the targeting of social media users, 13 April 2021; Blasi-Casagran, Vermeulen, 'Reflections on the murky legal practices of political micro-targeting from a GDPR perspective', 2021, 11 International Data Privacy Law 348; on the problem again Borgesius (fn. 21) and Towfigh (fn. 21).

[58] On the concerted use of these information claims, see for example: https://fragdenstaat.de/kampagnen/.

[59] Case C- 434/16 *Nowak,* 2017, paras 27 et seq.; OVG Münster, judgement of 8 June 2021 - 16 A 1582/20 (juris); see for instance the campaign on examination offices at https://fragdenstaat.de/kampagnen/.

[60] Case C-272/19 *VQ v Hesse,* 2020, paras 63 et seq.

[61] Since 2015, the EastStratCom Task Force of the European External Action Service has been collecting relevant material and publicly flagging disinformation on the site www.euvsdisinfo.eu and on social media. The basis for this is initially the conclusions of the European Council on External Relations of 19 March 2015, https://www.consilium.europa.eu/de/press/press-releases/2015/03/19/conclusions-russia-ukraine-european-council-march-2015/ and now the Action Plan against Disinformation of the High Representative for Foreign Affairs and Security Policy of 5 December 2018, JOIN, 2018, 36 final.

a) Harmful Content

The definition of illegal expressions of opinion, in particular the prohibition of hate speech or disinformation, continues to be a matter for the member states, without the European Union having any fundamental powers here. For this reason, the Union, based on competences regarding the internal market, criminal law, and the protection of democracy in the EU,[62] is proceeding on the one hand with soft law and on the other hand with selective legislation. As part of the Commission's general strategies against hate crime[63] and disinformation,[64] the Commission agreed in 2016 with the major digital platforms on a code of conduct for combating illegal hate speech online[65] and in 2018 on a (2022 reinforced) code of conduct for combating disinformation.[66] The self-obligations of the platforms, for example, to have procedures in place to review and remove illegal hate speech and disinformation, or to label political advertising, are in part a precursor to the platform regulation in the Digital Services Act (to be outlined below). In any case, self-regulation provides sufficient reason to define hate speech and disinformation at the European level as well. The Code of Conduct on Hate Speech refers to a Framework Decision, which has already defined racist and xenophobic speech offences in 2008, which are to be punished by the member states.[67] For disinformation, on the other hand, the corresponding code of conduct develops its own definition, which is based on two elements – the subjective element of economic gain or deliberate deception of the public, first, and the potential public damage, especially to democratic processes, health, environment, or security, second.[68] In addition, some varieties of disinformation – or techniques that enable them – are further specified in the code, such as the artificial amplification of information, deception with social bots and deepfakes, or the hijacking of user accounts.[69] Overall, the codes demonstrate that the

[62] See the European Action Plan for Democracy of the Commission of 3 December 2010, COM/2020/790 final; in addition, the competence for elections to the EU Parliament, in particular Art. 2, 3, 6 and Art. 9-21 Treaty on European Union can be used.

[63] Overview here: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination-0/racism-and-xenophobia/combating-hate-speech-and-hate-crime_en.

[64] Overview here: https://digital-strategy.ec.europa.eu/de/policies/online-disinformation.

[65] Available here: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination-0/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_de. The agreement was signed by Facebook, Microsoft, Twitter, and YouTube.

[66] Available here: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation. The agreement was signed by Facebook, Google, Twitter, Mozilla, Microsoft, TikTok, and other companies.

[67] Art. 1 Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.

[68] Preamble of the EU Code of Conduct on Combating Disinformation, 2018.

[69] See Commitment 14 of the Strengthened Code of Practice on Disinformation, 2022.

European Union considers it desirable for platforms to take action against harmful content (and the aforementioned harmful techniques) and does not see this as a basic problem for the fundamental rights of users or platforms.

Finally, under the Terrorist Content Regulation,[70] platforms are also already legally obliged to take selective action against prohibited content. The Regulation, based on internal market competence, refers to the offences under the Counter-Terrorism Directive[71] and defines terrorist content as, among other things, material that incites, glorifies, or avocates the commission of terrorist offences.[72] If national authorities discover terrorist content and issue a removal order, platforms are obliged to remove it as soon as possible and in any case within one hour.[73]

b) Problematic Communication Techniques

In addition, the Union legislature has also turned its attention to some harmful communication technologies. First of all, the planned AI Act should be mentioned here, which is based on the competences for the internal market and data protection.[74] The draft AI Act primarily regulates so-called high-risk applications of artificial intelligence, such as profiling and risk analysis in the areas of employment, lending, social benefits, or security.[75] In addition, Art. 52 of the AI Act establishes general transparency obligations for all AI systems. Accordingly, chatbots, i.e., computer programmes for interaction with humans, must be developed by their providers[76] in such a way that they are recognisable to the human user as a bot.[77] Not only service bots for customer service, but also social bots for the dissemination of political news must therefore be labelled as such and corresponding deceptions about a human interlocutor must be prevented.

[70] Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (for hosting providers).

[71] Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism; this in turn is based on the criminal law competence title under Art. 83 Treaty on the Functioning of the European Union.

[72] Art. 2 no. 6 and 7 Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online.

[73] Art. 3 para 3 Regulation (EU) 2021/784.

[74] Critical of the question of whether the regulations are covered by internal market competence, Valta, Vasel, 'Kommissionsvorschlag für eine Verordnung über Künstliche Intelligenz', ZRP 2021, 142, 143 et seq.; Linardatos, 'Auf dem Weg zu einer europäischen KI-Verordnung - ein (kritischer) Blick auf den aktuellen Kommissionsentwurf', 2022, GPR 58, 58 et seq.

[75] S. annex III and Art. 6 et seq. Artificial Intelligence Act (Proposal of 21 April 2021).

[76] Legal definition of 'provider' in Art. 3 no. 2 Artificial Intelligence Act (Proposal of 21 April 2021).

[77] Art. 52 para 1 sentence 1 Artificial Intelligence Act (Proposal of 21 April 2021).

Furthermore, the AI Act also obliges transparency when users[78] use deceptive deepfakes, unless freedom of speech or art requires an exception to the necessary disclosure.[79] Finally, another transparency requirement concerns the use of emotion recognition systems.[80] In addition, Art. 5 of the AI Act prohibits the marketing, commissioning, and use of some particularly harmful AI systems. This prohibition covers systems of behavioural manipulation in which the recipients are influenced subliminally, i.e., outside of awareness, or their particular weaknesses (age, disability) are exploited – with the aim of substantially influencing behaviour that causes or may cause psychological or physical harm to persons.[81] In the political context, this can be the case if the addressees are incited to violence with the help of these techniques. These provisions of the AI Act – despite all the criticism that can be voiced in view of imprecise or limited regulations[82] – are an important overall element in preventing deception and manipulation in online communication.

Another important project is the planned EU Regulation on political advertising.[83] Of all the legal acts mentioned so far, this regulation relates most strongly to political communication alone. It regulates political advertising, i.e., the elaboration and dissemination of a message that is made by or for a political actor or that is likely to influence a political decision (election, referendum, regulation).[84] The regulation is not limited to political actors or decisions at the EU level, but also covers domestic situations.[85] In terms of content, the regulation, based on internal market competence, first requires transparency from the providers of political advertising services. With these standards,

[78] Legal definition of 'users' as persons and entities using AI systems on their own authority and not merely in the course of a personal and non-professional activity, in Art. 3 no. 4 Artificial Intelligence Act (Proposal of 21 April 2021).

[79] Art. 52 para 3 Artificial Intelligence Act (Proposal of 21 April 2021); a further exception exists in favour of combating criminal offences.

[80] Art. 52 para 2 Artificial Intelligence Act (Proposal of 21 April 2021).

[81] Art. 5 para 1 lit a and b Artificial Intelligence Act (Proposal of 21 April 2021).

[82] Critical appraisal in Ebert, Spiecker gen. Döhmann, 'Der Kommissionsentwurf für eine KI-Verordnung der EU', NVwZ 2021, 1188, 1189 and 1191: inaccurate; Rostalski, Weiss, 'Der KI-Verordnungswurf der Europäischen Kommission', ZfDR 2021, 329, 337 et seq.; 350 et seq.: too narrow; Veale, Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' 22, 2021, Cri 97, 100 and 106 et seq.: lack of effectiveness; Kumkar, Rapp (fn. 22), 224 et seq.: inaccurate.

[83] Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (Political Advertising Regulation), 25 November 2021, COM/2021/731 final.

[84] Art. 2 no. 2 Political Advertising Regulation (Proposal of 25 November 2021): 'For the purposes of this Regulation, the term [...] 'political advertising' means the preparation, placement, promotion, publication or dissemination of a message by any means: (a) by or for, or on behalf of, a political actor, unless it is of a purely private or purely commercial nature; or (b) which is likely to influence the outcome of an election or referendum, a legislative or regulatory process or a vote'.

[85] Art. 2 no. 2 and 4 Political Advertising Regulation (Proposal of 25 November 2021).

the legislature wants to create "more legal certainty" for service providers without curtailing national competence for substantive issues of political opinion-forming (think, for example, of the limits of freedom of expression, election campaign rules, or party bans).[86] Transparency means, in particular, that political advertising must be labelled as such and with further information, such as the identity of the sponsor, the political context, and the aggregated costs.[87] Furthermore, the regulation, additionally based on the legislative powers concerning data protection, establishes rules for advertising that is shown only to certain persons or groups of persons (targeting) or whose reach or visibility is increased (amplification).[88] These regulations clearly go beyond existing approaches, for example at the Union level for advertising in audiovisual media[89] or at the national level in the State Media Treaty.[90] Meanwhile, the Digital Services Act also obliges platforms to be transparent about commercial and non-commercial online advertising.[91] However, the planned regulation retains its importance because it is primarily aimed at the advertising industry and political clients and also takes into account the special features of political communication. While the DSA prohibits advertising based on profiling using sensitive categories of data under Art. 9 GDPR,[92] the Regulation on political advertising takes into account the legitimate need of political parties to tailor advertising precisely to their members and interested parties – i.e., based on the sensitive datum of a political opinion.[93] In addition, further transparency requirements apply to the permitted constellations of targeting and amplification, so that the addressees can understand, in particular, the underlying logic and the most important parameters of the procedures used.[94]

---

[86] P. 6 et seq., Political Advertising Regulation (Proposal of 25 November 2021).

[87] Art. 7 p. 6 et seq., Political Advertising Regulation (Proposal of 25 November 2021).

[88] On terminology Art. 2 no. 8 Political Advertising Regulation (Proposal of 25 November 2021); data protection competence refers to the use of personal data, especially in targeting, see p. 7.

[89] Transparency of commercial communication and prohibition of subliminal influence according to Art. 9 para 1 Directive 2010/13/EU (Audiovisual Media Services Directive, consolidated version of 18 December 2018).

[90] Advertising must therefore be clearly recognisable and distinguishable from other content, and no techniques of subliminal influence may be used, see § 8 para 3, § 22 para 1 MStV. If political advertising (in telemedia) is permitted, 'the advertiser or client must be clearly indicated in an appropriate manner', § 22 para 1 sentence 3 MStV. In addition, social bots must be identified by telemedia providers in social networks, § 18 para 3 MStV.

[91] Art. 26 and Art. 39 DSA.

[92] Art. 26 para 3 DSA.

[93] Art. 12 Political Advertising Regulation (Proposal of 25 November 2021) prohibits the use of sensitive personal data under Art. 9 GDPR for targeting or amplification purposes altogether, unless explicit consent is given or unless parties and political associations communicate with their members and other existing contacts (Art. 9 (2) (a) and (d) GDPR).

[94] Art. 12 para 3-8 in conjunction with annex II Political-Advertising-Regulation (Proposal of 25 November 2021) .

3. The Digital Services Act

Going beyond the aforementioned special regulations, the EU is now focusing on platforms as key players in the online world with the Digital Services Act (DSA), based on the internal market competence of Art. 114 TFEU. For the sake of conceptual clarity, it should be emphasised that the DSA uses a narrower concept of platform than this article, while a large part of the intermediary services regulated by the DSA can be regarded as platforms in the broader sense. The DSA regulates firstly intermediary services (service providers),[95] secondly the subset of hosting services, i.e., intermediary services where information provided by a user is stored on behalf of that user,[96] and thirdly the subset of "online platforms", i.e., hosting services whose essential function is to store and publicly disseminate information on behalf of a user.[97] (This is to be distinguished from online search engines, which are also covered by the DSA.[98]) Fourthly and finally, the DSA lays down special requirements for "very large" online platforms (VLOPs for short) and search engines with an average monthly number of at least 45 million active users in the Union.[99] Online platforms in the sense of the DSA are thus, in particular, social networks such as *YouTube*, *Facebook* or *Twitter,* but not search engines, probably not messenger services (*WhatsApp*, *Telegram*),[100] and only partially trading platforms and streaming portals.[101] Instead of an overview of the comprehensive set of rules of the DSA, this article will present three central legal issues.

a) Platforms as Standard Setters

A first question concerns the freedom of platforms to set their own communication standards for their digital communication space. Platforms such as *YouTube*, *Facebook,* or *Twitter* are operated by companies under private law, which can set communication standards for their users on the basis of their private autonomy, for example, to promote a certain debate culture or to protect users from hate speech or disinformation. The corridor of permissible online content is then limited not only by state laws but also by

[95] Art. 3 lit. g DSA.
[96] Art. 3 lit. g (iii) DSA.
[97] Art. 3 lit. i DSA.
[98] Art. 3 lit. j DSA.
[99] Art. 33 para 1 DSA.
[100] In any case, messenger services qualify as hosting services. In order to be classified as an online platform, it would be necessary that their main task is to publicly disseminate user information (see Art. 3 lit. k DSA), which is doubtful, since they primarily enable communication between individuals and in closed groups.
[101] Intermediary services mediate users' content. Trading platforms and streaming portals are therefore only covered by the DSA to the extent that they distribute content of their users and not their own offers (e.g., Amazon as a sales platform for third parties, but not for Amazon itself; Spotify as a platform for third party podcasts, but not for Spotify's own productions).

these private-law guidelines. The question here would therefore be whether the platforms' power to set private standards is limited by fundamental rights such as the users' freedom of expression or other important legal interests. In German law, this is a case of indirect third-party effect. The Federal Constitutional Court first established criteria in the stadium ban decision that speak for an indirect third-party effect of the principle of equality vis-à-vis private parties: that private parties delibaretely open an event to a large audience without regard to the person, that an event is important for the participation in social life, or that a company has a strong position, in the sense of a monopoly or structural superiority.[102] Although these criteria were developed for German football, they are easily transferred to large platforms such as *Facebook.*[103] Here, the concrete drawing of boundaries ist crucial. The Federal Supreme Court recently affirmed that the platforms' private autonomy in principle also includes the right to restrict the users' freedom of expression[104] – specifically by banning hate speech – but that sanctions may only be imposed under certain procedural precautions (information, justification, possibility of counterstatement, and correction of the decision).[105]

For the DSA, there are indications of a comparable balance of interests. On the one hand, the DSA distinguishes between unlawful communications and communications in violation of platform rules. The DSA recognises the autonomy of service providers to set their own communication standards in their general terms and conditions, which can be used to restrict the users' activities.[106] On the other hand, the DSA explicitly obliges platforms to respect the fundamental rights of the European Charter of Fundamental Rights. Service providers are required to be "diligent, objective and proportionate" in the application and enforcement of their T&Cs and to take into account the rights and legitimate interests of all stakeholders as well as the fundamental rights of users, namely freedom of expression and freedom and plurality of the media.[107] In addition, VLOPs are required to assess and mitigate the systemic risks posed by their activities, in particular taking into account any adverse impact on fundamental rights in their deliberations.[108] These obligations also affect the scope of private autonomy.

[102] BVerfGE 2018, 148, 267, 284 et seq. *Stadionverbot*.

[103] BVerfG, NJW 2019, 1935, 1936 *III. Weg*.

[104] In some literature and case law this is seen more strictly, see for example the references in BGH NJW 2021, 3179, 3185.

[105] BGH NJW 2021, 3179, 3185 et seq.; see for example Raue, 'Plattformnutzungsverträge im Lichte der gesteigerten Grundrechtsbindung marktstarker sozialer Netze', NJW 2022, 209.

[106] Art. 14 para 1 DSA formulates mandatory requirements for the T&Cs of service providers, but at the same time affirms the right to the restrictions mentioned. On the conceptual difference between 'illegal content' and content that is 'incompatible' with the service providers' terms of use, see for example Art. 3 lit. h, Art. 6 para 1, Art. 9 para 1, Art. 16 para 1, Art. 17 para 1, Art. 20 para 1 DSA.

[107] Art. 14 para 4 DSA.

[108] Art. 34 para 1 lit. b, Art. 35 para 1 DSA.

Overall, therefore, the validity of fundamental rights in the private law relationship between platform and users follows from the DSA – irrespective of the question of the extent to which the fundamental rights of the Charter of Fundamental Rights have a general third-party effect.[109]

b) Platforms as a Regulatory Power for State and Private Law

A second question touches on the responsibility of platforms to combat illegal or anti-platform activities of third parties – the users – on their platform. This is also referred to as dealing with unlawful or anti-platform "content", which includes modalities of content delivery (e.g., deceptions) that violate state law or platform standards. For unlawful activities, it was initially the case under the E-Commerce Directive that platforms were essentially exempt from liability if they had no knowledge of unlawful content or acted expeditiously to remove or disable access to it upon becoming aware of it.[110] This liability privilege facilitated the desired activity of the platforms but did not lead to effective action against, for example, criminal offences of expression in the online space. Therefore, in 2017, the German legislature moved forward with the Network Enforcement Act (NetzDG). In particular, the NetzDG requires platforms to react to complaints about certain illegal content in a specific and refined[111] procedure, i.e., to remove or block specifically illegal content as a rule within seven days (in the case of obvious illegality, within 24 hours) as well as to justify this decision and to review and revise it after a cross-appeal.[112] The DSA now establishes comparable and significantly more far-reaching obligations at the European level. In this context, objections that were voiced about the NetzDG can also be transferred to the DSA. On the one hand, it is feared that platforms will act in a risk-averse and cost-saving manner and will therefore tend to classify content as illegal in case of doubt. This raises the issue of freedom-hostile "overblocking", which the platforms would be induced to do by the law.[113]

[109] On the state of the discussion, for example, Kingreen, in Calliess, Ruffert (eds.), EUV/AEUV, 6th ed. 2022, Art. 51 GRCh, para 24 et seq.; Schwerdtfeger, in Meyer, Hölscheidt (eds.), Charter of Fundamental Rights of the European Union, 5th ed. 2019, Art. 51 para 57 et seq.

[110] This obligation applies to all hosting providers, Art. 14 para 1 Directive 2000/31/EC (E-Commerce Directive).

[111] Cornils, 'Präzisierung, Vervollständigung und Erweiterung: Die Änderungen des Netzwerkdurchsetzungsgesetzes', NJW 2021, 2465.

[112] This obligation applies to social networks, § 3 para 1 and para 2 sentence 1 no. 2, 3 and 5, § 3 b NetzDG.

[113] Also on other points of criticism of the NetzDG (doubts about the federal government's legislative competence under EU and constitutional law, disproportionate burden on platforms) comprehensively Ladeur and Gostomzyk, 'Gutachten zur Verfassungsmäßigkeit des NetzDG', May 2017, https://www.cr-online.de/NetzDG-Gutachten-Gostomzyk-Ladeur.pdf; on the problem of overblocking, for example, Guggenberger, 'Das Netzwerkdurchsetzungsgesetz in der Anwendung', NJW 2017, 2577, 2581 et seq.;

On the other hand, one could criticise the fact that the platforms are now assigned not only the de facto but also the legal key position for order in the digital space, while law enforcement should be the responsibility of state authorities and courts – also on the internet.[114] Incidentally, both points of criticism are also justified when platforms (merely) enforce their own standards. Here, too, platforms act as a private regulatory power in the digital space and could – due to their orientation towards economic interests – choose simple solutions at the expense of fundamental rights.

The EU legislature has sought to strike an appropriate balance between these interests. This applies first of all to the effective fight against illegal content. The DSA explicitly does not impose a general obligation on platforms to monitor the information they transmit or store or to actively search for illegal content.[115] This corresponds to the ideal of a free society in which people live and communicate with each other in physical and digital spaces without surveillance, even if full surveillance now seems technically possible.[116] At the same time, the liability privilege of the platforms remains intact if they are ignorant of illegal content.[117] However, numerous accompanying norms are intended to bring about more effective action against illegal content. For example, the platforms are obliged to set up an easily accessible and user-friendly reporting and redress procedure for illegal content and to process tips "promptly".[118] The DSA also relies on the institute of recognised "trusted flaggers" for priority and immediate treatment of tips.[119] In the case of frequent and obviously illegal content, the DSA even requires the platforms to temporarily block a user from the platform.[120] Furthermore, the

Heckmann and Wimmers, 'Stellungnahme der DGRI zum Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (NetzDG)', CR 2017, 310, 314 et seq.; Müller-Franken, 'Netzwerkdurchsetzungsgesetz: Selbstbehauptung des Rechts oder erster Schritt in die selbstregulierte Vorzensur?', AfP 2018, 1, 1 et seq.; Hoven and Gersdorf, in Gersdorf, Paal (eds.), BeckOK Informations- und Medienrecht, 38th ed. 1 Mai 2021, § 1 NetzDG, para 6.

[114] Heckmann, Wimmers (fn. 113), 313 et seq.; Wimmers and Heymann, 'Zum Referentenentwurf eines Netzwerkdurchsetzungsgesetzes (NetzDG) - eine kritische Stellungnahme', AfP 2017, 93, 99; on alternatives therefore Guggenberger, 'Das Netzwerkdurchsetzungsgesetz - schön gedacht, schlecht gemacht', ZRP 2017, 98, 101.

[115] Art. 8 DSA; see previously Art. 15 E-Commerce Directive.

[116] On this ideal see, for example, from the case law on data protection, ECJ Case C-293/12 *Digital Rights Ireland*, 2014, paras 57 et seq. and the subsequent case law; also Art. 17 para 8 Directive (EU) 2019/790 (Directive on Copyright in the Digital Single Market).

[117] Art. 6 para 1 DSA (for hosting providers).

[118] Art. 16 DSA (for hosting providers).

[119] Art. 22 DSA. Such whistleblowers who have frequently attracted attention through obviously unfounded reports shall be temporarily excluded from the reporting system by the platforms, Art. 23 para 2 DSA.

[120] Art. 23 para 1 DSA 'for a reasonable period after prior warning' (for online platforms).

DSA proceduralises the cooperation of platforms and authorities[121] and obliges platforms to report suspicions of serious crimes against the life or safety of people.[122] In addition, VLOPs have to communicate and minimise systemic risks of the dissemination of unlawful content.[123]

Secondly, the DSA also aims at balancing conflicting interests in a concrete conflict about illegal or platform-infringing content. It is important to note that this conflict usually involves at least three parties: firstly, the platforms (which can rely on freedom of enterprise), secondly, the affected parties or whistleblowers who want to have the content removed (and can, for example, invoke personal rights or the protection of democratic processes for this purpose), and thirdly, those users who want to disseminate or view the content (based on freedom of expression or freedom of information).[124] Here, the DSA first orders that all measures against illegal or platform-unfriendly content must be justified, not only the removal or blocking of accounts or content but also a reduction in visibility and reach ("shadow banning").[125] Furthermore, online platforms must have an internal complaints management and an out-of-court dispute resolution system for users and whistleblowers in case of conflict.[126] It will continue to be the task of state jurisdiction to clarify fundamental questions about the scope of fundamental rights in the triangle of interests of the parties involved. However, the procedure envisaged by the DSA allows for appropriate handling of the mass conflicts by the platform itself with the participation of the relevant parties. The DSA recognises the need for the platforms to automate procedures to deal with large numbers of cases. In the moderation of content,[127] in the reporting and redress procedure,[128] and in the decision on measures,[129] the use of automation is permissible and must be disclosed; in complaints management, a decision may also be made under human supervision with the help of automated means.[130]

c) Platforms as a Black Box

This leads to a third legal question: the control of platforms. As private companies, platforms do not have to disclose their operating procedures and internal decisions –

[121] In the case of official orders - issued under national law - to take action against illegal content (Art. 9 DSA) or to provide information about users (Art. 10 DSA); applies in each case to service providers.
[122] Art. 18 para 1 DSA.
[123] Art. 34 para 1 sentence 3 lit b, Art. 35 DSA.
[124] On the latter in particular Raue, 'Meinungsfreiheit in sozialen Netzwerken', JZ 2018, 961.
[125] Art. 17 DSA (for hosting providers).
[126] Art. 20 and Art. 21 DSA (for online platforms).
[127] Art. 14 para 1, Art. 15 para 1 lit c DSA (for service providers).
[128] Art. 16 para 6 sentence 2 DSA.
[129] Art. 17 para 3 lit c DSA.
[130] Art. 20 para 6 DSA.

including the algorithms used. However, as indispensable forums for online communication, platforms are so central to the economy, society, and politics that there is a public interest in being able to understand and control their activities. This is especially true where the platforms use algorithms whose effects are difficult to gauge,[131] or where they act as a private regulatory power for online discourse. The DSA essentially relies on transparency here. Platforms have to publish regular reports and fan out their content moderation – with specific information both on moderation on their own initiative and on the procedure in the case of official orders, in the reporting and redress procedure, in internal complaint management, and in extrajudicial dispute resolution.[132] The relevant algorithms are to be described in more detail, with a precise statement of purpose, indicators of accuracy, possible error rate, and safeguards.[133] Special rules also apply to algorithmic recommendation systems that suggest or prioritise content for users[134] and thus essentially determine what users see online. Here, transparency includes explaining which parameters are most important for the recommendation system (e.g., characteristics, preferences, and networks of the user; respectability or popularity of a source; trends, etc.) and how these parameters can be changed or influenced by the user if necessary.[135] VLOPs must also offer their users the possibility to opt for a recommendation system that does not require profiling.[136] Finally, VLOPs are generally obliged to explain the design, logic, functioning, and testing of their algorithmic systems upon request.[137]

Regardless of this transparency, it may still be difficult to determine the positive or negative effects of algorithms. To what extent does the focus on maximum user interaction actually lead to a displacement of sober content by lurid, polarising, and extremist messages or to a parcelling of the digital space into small echo chambers? To what extent do deliberate moderation rules (such as the reduced reach of videos with obese or disabled people on TikTok)[138] or learning processes of intelligent algorithms (such

[131] For the common designation of algorithms as black boxes, see Pasquale, The Black Box Society, 2016.

[132] Art. 15 para 1 DSA (for service providers), Art. 24 para 1 DSA (for online platforms); cf. already the reporting obligations under § 2 NetzDG.

[133] Art. 15 para 1 lit e DSA.

[134] Art. 3 lit s DSA.

[135] Art. 27 para 1 and 2 DSA (for online platforms); a similar transparency requirement for media intermediaries is already found in section 93 of the State Media Treaty.

[136] Art. 38 DSA.

[137] Art. 40 para 3 DSA.

[138] The company's reasoning is that these people would otherwise easily become targets of cyberbullying, Köver, Reuter (fn. 35).

as on the preferences of online users) reinforce[139] stereotypes and discrimination?[140] And also, to what extent do the platforms succeed in automatically detecting problematic content and neutralising it? It is certainly gratifying when *Facebook* or *Twitter* takes action against Russian disinformation campaigns and also regularly publishes how many accounts and pages have been deactivated.[141] However, it is not possible to check whether the platforms were correct in doing so and which activities the companies missed. The DSA responds to this problem on the one hand with an obligation for VLOPs to analyse and mitigate systemic risks precisely from the use of algorithms.[142] As relevant risks, the DSA mentions, for example, the dissemination of illegal content, impairment of fundamental rights, and adverse effects on elections or public security.[143] On the other hand, the DSA opens the VLOPs to external control – not only by the official supervisory authority[144] or independent auditors,[145] but also by the scientific community. In particular, VLOPs are required to maintain a publicly accessible and searchable archive of online advertising conducted.[146] In addition, researchers will in the future have access to the data of VLOPs in order to be able to control how the platforms fulfil the task of risk analysis and risk minimisation.[147]

IV. Europeanisation of Freedom of Expression

The above-mentioned legislative projects involve a Europeanisation of freedom of expression in EU legislation (I), which will also be reflected in an increased responsibility of the ECJ for freedom of expression (II).

[139] Sweeney, 'Discrimination in online ad delivery', Commun. ACM 2013, 44, on specific advertising offers (information about criminal records) in Google searches on 'black' sounding names.

[140] On this, generally, von Ungern-Sternberg, 'Diskriminierungsschutz bei algorithmenbasierten Entscheidungen', in Mangold, Payandeh (eds.), Handbuch Antidiskriminierungsrecht – Strukturen, Rechtsfiguren und Konzepte, 2022, § 18 paras 10 et seq.

[141] Brandt, 'Facebook removes 52 manipulation networks', Statista, 21 January 2022, https://de.statista.com/infografik/24653/aktionen-von-facebook-im-zusammenhang-mit-coordinated-inauthentic-behavior/.

[142] Art. 34 para 1 sentence 1, para 2 lit a and sentence 2, Art. 35 para 1 sentence 2 lit d DSA.

[143] Art. 34 para 1 sentence 3 lit a-c DSA.

[144] Data access according to Art. 40 para 1 DSA.

[145] Art. 37 DSA.

[146] Art. 39 DSA.

[147] Art. 40 para 4 DSA.

1. EU Legislation on Freedom of Expression

The overview of European legislation shows: the laws shaping the digital space, especially laws on the digital economy and data protection, also provide the framework for freedom of expression. It is true that the definition of illegal content remains predominantly in the hands of the member states. But the Union legislature determines how platforms deal with illegal and platform-unfriendly content: it tackles the systemic risks of VLOPS for the democratic formation of opinion, it restricts some particularly problematic techniques of online research, deception, and manipulation, and it shapes the relationship between data protection and freedom of expression. Private parties, especially platforms or data processors, are explicitly obliged to respect freedom of expression or fundamental rights as a whole.[148] In addition, the obligation of state and EU actors to protect fundamental rights, especially freedom of expression (which already exists according to Art. 51 (1) sentence 1 CFR) is partly reaffirmed,[149] a Member State's leeway to protect freedom of expression and media is recognised,[150] and freedom of expression forms a limit for legislative measures.[151] Finally, the recitals of the legislative proposals underline that the Union legislature aims to protect fundamental rights and, in particular, freedom of expression.[152] Consequently, the EU legislature does not see its activity as purely economic, but rather, in the broader context of the European Action Plan for Democracy, pursues the goal of "building more resilient democracies across the Union" (according to the explanatory memorandum to the DSA[153]) or "protecting the integrity of elections and promoting democratic participation" (according to the explanatory memorandum to the Political Advertising Directive).[154]

This Europeanisation is to be welcomed. One can hardly entrust the challenges to democratic opinion-forming to the platforms alone. Even if they seem to be trying to

---

[148] See again Art. 14 para 4, Art. 34 para 1 lit b, Art. 35 para 1 sentence 1 DSA (platforms); furthermore Art. 5 para 1 upara 2, para 3 upara 1 lit c Regulation (EU) 2021/784 (Regulation on terrorist content) (hosting provider); Art. 6 para 1 sentence 1 lit f GDPR (data processing based on legitimate interest), Art. 22 para 3 GDPR (automated decision-making).

[149] Art. 1 para 1 lit b Regulation on terrorist content.

[150] Art. 85 para 1 and 2 GDPR.

[151] For example, the prohibition of deepfakes, Art. 52 para 3 upara 2 Artificial Intelligence Act (Proposal of 21 April 2021); as the limit of the right to erasure under data protection law Art. 17 para 3 lit a GDPR.

[152] See Recitals 3, 9, 22, 36, 40, 41, 47, 51-54, 63, 79, 81, 86, 90, 107, 109, 116, 149, 150, 153, 155 DSA, also for example Art. 1 para 1 DSA; Recitals 12, 31, 47, 58 Political Advertising Regulation (Proposal of 25 November 2021); Recitals 10, 49 Regulation (EU) 2021/784 (Regulation on terrorist content); Recital 28 Artificial Intelligence Act (Proposal of 21 April 2021); Recitals 4, 65, 153 DSA.

[153] Explanatory Memorandum to the Commission Draft DSA, 15 December 2020, COM/2020/825 final, p. 5.

[154] Political Advertising Regulation (Proposal of 25 November 2021), Explanatory Memorandum, p. 5.

combat harmful content and techniques after the revelations about Russian disinformation, the Cambridge Analytica scandal, or the storming of the Capitol, they are first and foremost committed to their business interests. It is therefore logical and appropriate in the systemic competition with the USA and China that Europe-wide legislation should be adopted in the European single market.[155]

2. EU Case Law on Freedom of Expression

This also means that the ECJ will deal more intensively with freedom of expression and information as protected by Art. 11 CFR. In Germany, the ECJ's case law occasionally provokes defensive reflexes. When the president of the ECJ, Koen Lenaerts, was once asked by the FAZ to what extent his court was to blame for the crisis in the EU, he replied: "You notice right away that you are in Germany".[156] On sober reflection, however, the fear of a lack of institutional protection of fundamental rights before the ECJ[157] has been dispelled from a German perspective with the Federal Constitutional Court's decisions on the right to be forgotten. This means that the review and application of EU fundamental rights is no longer confined to the ordinary German courts (and the ECJ, at least by way of preliminary rulings), it is also a matter for the Federal Constitutional Court in the context of a constitutional complaint.[158] The concern remains that the ECJ might accord too little importance to freedom of expression, especially when weighing it against other fundamental rights.[159] After all, in the *Google Spain* decision on the balancing of data protection and freedom of expression (or freedom of information), the problematice sentence can be found that the data subjects' rights to data protection "override, as a rule, not only the economic interest of the operator of the search engine but also the interest of the general public in finding that information upon a search relating to the data subject's name".[160] In the following, however, it should be argued that, on the basis of previous case law, there is no reason to fear that the ECJ might neglect freedom of expression.

a) Rudiments of Jurisprudence on Freedom of Expression

For a long time, freedom of expression – initially as an unwritten legal principle and since 2009 guaranteed under Art. 11 CFR – hardly played a role in the case law of the

[155] The fact that the DSA (similar to the GDPR) is likely to impose overly far-reaching bureaucratic requirements in some places will not be discussed in more detail here.

[156] Müller, 'President for EU Values', FAZ, 16 February 2022, https://www.faz.net/aktuell/politik/ein-praesident-fuer-die-werte-der-eu-17810540.html.

[157] See above fn. 9.

[158] *Recht auf Vergessen II,* BVerfGE 2019, 152, 216 (236 et seq.).

[159] This is also how the fear expressed in the FAZ can be understood, see above, fn. 2.

[160] Case C-131/12 *Google Spain*, 2014, para 97; see above. fn. 11.

ECJ.[161] The few decisions deal, for example, with public statements by EU officials,[162] the rights of EU parliamentarians,[163] or the limits of competition and free movement in favour of freedom of expression.[164] This is not surprising as long as EU law had no corresponding links to freedom of expression. Especially democratic opinion making was considered a national matter, just like the choice of the governmental system or party and electoral law.

However, the European Convention on Human Rights lays down requirements that have a direct impact on democratic decision-making. For example, the European Court of Human Rights has ruled on the right of prisoners to vote,[165] on the participation of Gibraltarians in the elections of the European Parliament,[166] and on the prerequisites for party prohibition proceedings.[167] Above all, however, the ECtHR has developed extensive case law on freedom of expression as protected by Art. 10 ECHR. The ECJ has to consider and follow this case law,[168] which it in fact does. The ECtHR's case law on Art. 10 ECHR resembles or corresponds in many cases to German case law on the scope of freedom of expression.[169] Occasionally, German law has been forced to

[161] Overview e.g., Cornils, in Gersdorf, Paal (eds.), BeckOK Informations- und Medienrecht, 6th ed. 1 February 2021, Art. 11 GRCh paras 1 et seq.

[162] Case C-100/88 *Oyowe and Traore v Commission*, 1989, para 16: Official's duty of loyalty does not conflict with freedom of expression.

[163] Case C-163/10 *Patriciello*, 2011, paras 29 et seq.: broad understanding of parliamentary immunity also for mandate-related statements outside parliament.

[164] Case 43 a. 63/82 *VBVB a. VBBB*, 1984, paras 33 et seq.: Action against book price cartel does not affect freedom of publication; Case C-260/89 *ERT*, 1991, para 44: Television monopoly may be justified, Art. 10 ECHR as a barrier to the freedom to provide services; Case C-368/95 *Familiapress*, 1997, paras 24 et seq.; prohibition of prize competitions in magazines can be justified, Art. 10 ECHR here concurrent with the freedom of movement of goods (cross-border information) and barrier (protection of media diversity); Case C-112/00 *Schmidberger*, 2003, paras 69 et seq.: state non-intervention against street blockades approved at the expense of the freedom of movement of goods to protect the freedom of assembly.

[165] Arndt, in Karpenstein, Mayer (eds.), EMRK, 3rd ed. Munich 2022, Art. 3 ZusProt EMRK, para 24; v. a.a. ECtHR (GC) No 74025/01, 6 October 2005, *Hirst/UK*; ECtHR (GC) No 126/05, 22 May 2012, *Scoppola v Italy* No 3.

[166] ECtHR (GC) No. 24883/94, 18 February 1999, *Matthews/UK*.

[167] Arndt, Engels, von Oettingen, in Karpenstein, Mayer (fn. 165), Art. 11 EMRK, para 42; in particular ECtHR (GC) No. 41340/98, 12 February 2003, *Refah Partisi/Turkey*.

[168] The coherence requirement of Art. 52 para 3 sentence 2 GRCh as well as the minimum level under convention law according to Art. 53 GRCh apply.

[169] Admittedly, there are numerous dogmatic differences, such as the application of Art. 10 ECHR to rights that are protected in Germany with specific fundamental rights (freedom of occupation, freedom of art), the abuse clause of Art. 17 ECHR (according to which the ECtHR denies protection to racist statements), or the construction of the rights' limitations according to Art. 10 para 2 ECHR and Art. 5 para 1 sentence 2 GG. But in terms of the material scope, the right to freedom of expression of the Basic Law and the ECHR should essentially differ in that the ECtHR respects a margin of appreciation of the member states. Moreover, the ECtHR and the Federal Constitutional Court both emphasise the functional dimension of freedom of expression for the protection of democracy, both courts protect freedom of expression

adapt. For example, the *von Hannover* case law of the ECtHR has led the German courts to weigh the freedom of the press against the personality rights of celebrities somewhat more strongly in favour of personality rights.[170] Even if one can certainly argue about individual accentuations of freedom of expression, there is nothing fundamental to object against the ECtHR's case law in this regard.

b) Strengthening Freedom of Expression in the Recent Case Law of the Court of Justice

More recently, the case law of the European Court of Justice has increasingly included decisions in which freedom of expression plays a decisive role. In particular, the ECJ uses freedom of expression to interpret secondary legislation, such as copyright law,[171] media law,[172] and data protection law.[173] This is a logical consequence of legislative activity in these areas, which directly affects the fundamental right to freedom of expression and information. At the same time, since the entry into force of the Lisbon Treaty, a stronger focus of case law on EU fundamental rights can be observed. In the following, three examples will outline how the Court of Justice strengthens freedom of expression even in the face of conflicting interests.

First of all, data protection should be considered once again, since it is precisely the *Google Spain* case law that has caused the above-mentioned fears of insufficient protection of freedom of expression. Generally speaking, the ECJ does not set data protection in absolute terms but always takes freedom of expression or the public's interest in information into account as potential limits.[174] Furthermore, the ECJ also takes a broad

in private-law relationships and, when drawing the line, take into account a variety of factors such as the context of expression, the public interest in an expression and, in particular, the watchdog role of the media as well as the severity of a restriction. In detail Grote and Wenzel, in Dörr, Grote, Marauhn (eds.), EMRK/GG Konkordanzkommentar, 3rd ed. Munich 2022, Kapitel 18 Die Meinungsfreiheit, marginal nos. 10, 14, 17 et seq., 21 et seq., 32 et seq., 35, 57 et seq., 100 et seq., 105 et seq., 138. See also Mensching, in Karpenstein, Mayer (fn. 165), Art. 10 EMRK, marginal nos. 1 et seq., 52 et seq.

[170] ECtHR No 59320/00, 24 June 2004, *Hannover ;* ECtHR No 40660/08, 7 February 2012, *Hannover II*; see Hong, 'Caroline von Hannover und die Folgen' in Matz-Lück, Hong (eds.) Fundamental Rights and Freedoms in the Multi-level System, 2012, 251 et seq.

[171] Case C-70/10 *Scarlet Extended,* 2011, para 52: freedom of information protection against extensive filtering of electronic communications; also case C- 360/10 *SABAM*, 2012, para 50; also case C-145/10 *Painer*, 2011, paras 114 et seq.: Copyright protection also vis-à-vis the press ; case C-314/12 *UPC Telekabel Wien*, 2014, paras 55 et seq.: Freedom of information in the case of internet access blocks.

[172] Case C-283/11 *Sky Austria*, 2013, paras 51 et seq.: exclusive television reporting rights and short reporting by third parties in favour of freedom of information.

[173] Case C-73/07 *Markkinapörssi*, 2008, paras 52 et seq.: Journalist's privilege for publication of tax data; Case C-345/17 *Buivids*, 2019, paras 48 et seq.: Journalist's privilege for publication of video of police interrogation.

[174] More cautiously formulated (transparency interest of the Union) Case C-92/09 *Schecke*, 2010, paras 77; more strongly, for example, Case C-101/01 *Lindqvist*, 2003, paras 72 et seq. (on freedom of expression).

view of the media privilege and thus of the opening clause for the member states in this regard. It regards bloggers as so-called "citizen journalists" if their contribution to an online platform has the "sole purpose" of "disclosure to the public of information, opinions or ideas".[175] Finally, notwithstanding the unfortunate phrase of the *Google Spain* decision, it should be emphasised that the ECJ seeks an "appropriate balance" between the fundamental rights affected in the delisting cases,[176] that it establishes comprehensible criteria for a delisting (the sensitivity of the information, its topicality, any public importance of the person concerned[177]), and, above all, that the weighing in the concrete individual case remains the responsibility of the member state courts. This division of labour between the ECJ and the national courts, like the recognition of a certain "margin of appreciation" in the member states, is to be welcomed and is appropriate in view of democratic opinion-forming.[178]

A second example of the scope of freedom of expression concerns the freedom of speech of members of Parliament. According to the case law of the ECtHR and the ECJ, MEPs enjoy freedom of speech under Art. 10 ECHR and Art. 11 CFR, respectively, and can defend themselves on this basis against parliamentary disciplinary measures.[179] The ECJ recently had to review two disciplinary measures against a right-wing populist MEP who had been sanctioned for offensive (racist and sexist) statements. The ECJ upheld the actions because there was no serious breach of order or serious disruption of the work of the Parliament to justify the interference with freedom of speech.[180] Both decisions are certainly worthy of criticism in their outcome,[181] but they prove that the ECJ case law by no means has a general tendency to deny the protection of freedom of expression in cases of "hate speech".

Thirdly, the ECJ has also strengthened freedom of expression in the field of copyright. In particular, there is a dispute about whether platforms such as *YouTube* may or

---

[175] So Case C-345/17 *Buivids*, 2019, para 59; also Case C-73/07 *Markkinapörssi,* 2008, paras 52 et seq.

[176] Case C-131/12 *Google Spain*, 2014, para 81.

[177] Case C-131/12 *Google Spain*, 2014, further Case C-136/17 *GC and Others v CNIL,* 2019, paras 59 et seq., 75 et seq.

[178] On the recognition of margins of appreciation and discretion by the ECJ see for example Sweeney, 'A 'margin of appreciation' in the internal market: lessons from the European Court of Human Rights', Legal issues of economic integration 34 (2007), 27; Walter,Vordermeyer, 'Verfassungsidentität als Instrument richterlicher Selbstbeschränkung in transnationalen Integrationsprozessen', JöR 63 (2015), 129, 130 et seq.; Oreschnik, Verhältnismäßigkeit und Kontrolldichte, 2018; Schulte, Zur Übertragbarkeit der Margin-of-appreciation-Doktrin des EGMR auf die Rechtsprechung des EuGH im Bereich der Grundfreiheiten, 2018, 184 et seq.

[179] For a detailed discussion, see Kemper, Parlamentarische Redefreiheit im Spannungsverhältnis von Status- und Grundrechten, PhD thesis Trier, 2023, chapters 2 and 4, forthcoming.

[180] Case T-770/16 *Korwin-Mikke I*, 2018, para 79 ; Case T-352/17 *Korwin-Mikke II*, 2018, para 71.

[181] Sauer, 'Basic Rights as a Basis for Populism in Parliament?', Verfassungsblog 4 June 2018, https://verfassungsblog.de/grundrechte-als-grundlage-fuer-populismus-im-parlament/; Kemper (fn. 179), chapter 4.

should use upload filters to prevent copyright infringements, so that the mass uploading of protected works is prevented. The ECJ initially ruled that the use of upload filters was unlawful because it violated freedom of information.[182] In the meantime, the EU legislature has amended the law to the effect that not only the users who upload a protected work but also the platforms are liable for copyright infringement. However, according to the controversial Article 17 (4) of the DSM Directive, liability is excluded if the platform has made best efforts to obtain an authorization, and has "made, in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works".[183] This due diligence requirement can probably only be met by using upload filters.[184] Critics fear that, in order to avoid the risk of liability, platforms will make their filters so sharp that they will block even legal uploads in case of doubt ("overblocking"). A user would then not be able to upload any permitted quotes, parodies, or memes and would therefore be prevented from expressing her opinion. Moreover, critics see the danger of a precedent that would establish the instrument of upload filters and make it acceptable for other areas such as hate speech or disinformation.[185] Against this background, the German implementation of the DSM Directive takes a thoroughly fundamental-rights-friendly approach: It presumes that certain uses are permitted, which cannot be blocked in advance (for example, a film excerpt up to 15 seconds in length)[186] and, in the event of a blockade, provides for a complaints procedure involving all parties.[187] The ECJ has already had the opportunity to review the conformity of Art. 17 of the DSM Directive with fundamental rights at the request of Poland. The Grand Chamber of the ECJ first found that Article 17 (4) of the DSM Directive contains a restriction of the users' fundamental right to freedom of expression and information, even if the provision only grants an exemption from liability in legal terms.[188] Like the Advocate General, the Grand Chamber also affirmed the proportionality of the interference by interpreting the Directive in conformity with fundamental rights. Accordingly, platforms may only use upload filters where they function reliably

[182] See Case C-70/10 *Scarlet Extended*, 2011, para 52: protection of freedom of information against extensive filtering of electronic communications; also Case C- 360/10 *SABAM,* 2012, para 50.

[183] Art. 17 para 4 lit a and b Directive on Copyright in the Digital Single Market (DSM Directive).

[184] Müller-Terpitz, 'Filter als Gefahr für die Meinungspluralität?', ZUM 2020, 365, 366; Metzger, Pravemann, 'Der Entwurf des UrhDaG als Umsetzung von Art. 17 DSM-RL', ZUM 2021, 288, 292; AG Henrik Saugmandsgaard Øe, 5 July 2021, Case C-401/19 *Poland/EP and Council*, 2022, paras 68 et seq.; Case C-401/19 *Poland/EP and Council*, 2022, para 54.

[185] Horchert, 'Lasst euch das Internet kurz erklären, bevor ihr es kaputt macht', SPIEGEL, 23 March 2019, https://www.spiegel.de/netzwelt/netzpolitik/artikel-13-und-uploadfilter-zehntausende-protestieren-gegen-urheberrechtsreform-a-1259360.html.

[186] §§ 9 and 10 Act on the Copyright Responsibility of Service Providers for the Sharing of Online Content (UrhDaG), on this Metzger, Pravemann (fn. 184).

[187] §§ 14, 15, UrhDAG.

[188] Case C-401/19 *Poland v EP and Council,* 2022, paras 48 et seq.

so that blocking of lawful content during uploading is excluded.[189] In addition, the legislature is obliged to ensure that users may upload and share self-generated content for the purposes of quotations, criticism, reviews, caricatures, parodies, and pastiches, i.e., that these certain formats – protected by the freedom of art or expression – are not prevented by upload filters.[190] Furthermore, the ECJ emphasised that Art. 17 should not lead to a general monitoring obligation[191] and that there would still be procedures in which the platform could remove content only after being notified by a rights holder or only after independent substantive examination.[192] Finally, in order to affirm proportionality, the ECJ also stressed that users could effectively proceed against a content block by way of complaint or legal remedy.[193] The ECJ decision thus approved the legislative decision in favour of upload control in principle, but - like the German implementation - provided it with very significant restrictions to protect freedom of expression. Details, however, still have to be specified.[194]

V. Conclusion

The EU legislature not only regulates the digital economy but also shapes the space of digital opinion-making and the public sphere. The aim is to strengthen and consolidate European democracy(ies), and this proves that the European Union sees itself not only as an economic union but also as a political union with common fundamental values (Art. 2 TEU). Consequently, EU legislation also leads to a Europeanisation of freedom of expression, i.e., to the fact that the scope and limits of freedom of expression are now increasingly in the hands of the ECJ. This requires a critical assessment by legal scholars. However, there is nothing to suggest so far that freedom of expression is not in good hands with the ECJ. The ECJ might accentuate the weighting of freedom of expression and opposing legal interests differently than German scholars are used to in German law. But a European legal framework protecting democratic opinion making is worth it.

---

[189] AG Henrik Saugmandsgaard Øe, 5 July 2021, C-401/19 *Poland/EP and Council,* 2022, paras 305 et seq.; Case C-401/19 *Poland/EP and Council*, 2022, paras 85 et seq.

[190] Case C-401/19 *Poland v EP and Council*, 2022, paras 87 et seq.

[191] Art. 17 para 8 DSM Directive.

[192] Case C-401/19 *Poland v EP and Council*, 2022, paras 90 et seq.

[193] Case C-401/19 *Poland v EP and Council,* 2022, paras 93 et seq.

[194] Cf. Kraetzig, 'Lennartz, Grundrechtsschutz durch Algorithmus?', NJW 75 (2022), 2524 et seq.; Lennartz, 'Digitale Filter zwischen Konsum und Kommunikation', EuGRZ 2022, 482 et seq.; Raue, 'Die Zähmung der Uploadfilter', ZUM 2022, 624 et seq.; 'Spindler, Der Streit um Art. 17 DSM-Richtlinie – endgültige Klärung durch den EuGH?', Computer und Recht 2022, 444 et seq.

Taking or Escaping Legislative Responsibility? EU Fundamental Rights and Content Regulation under the DSA

Mattias Wendel

I. Introduction

The adoption of the Digital Services Act (DSA)[1] is a milestone in the regulation of digital services in Europe. It provides a detailed, though partly unclear, incomplete, and controversial legal framework for regulating digital intermediary services, including social media platforms. As such, it addresses complex issues of liability and consumer protection. One thing is certain. The DSA will keep lawyers busy for many years to come. The vast majority of the open questions that need to be answered are matters of civil law – presuming that the categorical distinction between civil law and public law is still of significance when it comes to shaping the law of the digital space. However, especially in the area of content regulation, the DSA also raises important issues of constitutional law, particularly with regard to the protection of fundamental rights.

Against this background, this paper focuses on the protection of European – and national – fundamental rights in the field of content regulation, paying particular attention to the responsibility of the legislature to predetermine the balancing of conflicting fundamental rights at least to a certain degree. The article does not undertake a detailed analysis of the individual provisions of the DSA, but attempts to place the fundamental rights issues raised by the new regime in the more general context of the system of fundamental rights protection in the EU, namely its pluralist structure.

In a first step, the pluralist structure of the European fundamental rights architecture will be outlined in more general terms, covering the plurality of fundamental rights levels, standards, and relations as well as the plurality in enforcement. Specific attention will be paid to the federal management of the overlapping spheres of European and

[1] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19. October 2022 on a Single Market For Digital Services. While several Articles have already entered into force on 16. November 2022, the majority of provisions shall apply from 17. February 2024 (cf. Article 93 DSA).

national fundamental rights and the conditions of the horizontal (direct) effect of fundamental rights under EU law. Illuminating these aspects provides the basis for a better understanding of the specific challenges of fundamental rights protection in the context of the DSA (see 2.).

In a second step, it will be examined how this pluralist structure plays out with specific regard to content regulation under the DSA. The paper analyses which levels of fundamental rights protection matter in the context of the DSA and which standards are relevant from the perspective of users, platforms, and other players. In particular, it will be assessed to what extent fundamental rights can also be – directly or indirectly – invoked against the platforms and whether the flexibility granted by the EU legislature to the platforms, but also to the courts, in weighing up the conflicting fundamental rights is constitutionally sound. Finally, the modalities of enforcement, in particular the judicial and extrajudicial mechanisms of dispute resolution, are addressed (see III.).

The paper concludes that, in terms of constitutional law, the decisive question seems not so much whether the relevant fundamental rights, such as freedom of expression and information, can have a direct horizontal effect and directly impose obligations on platforms, but rather to what extent the EU legislature must predetermine the balancing of conflicting fundamental rights and thus limit the leeway of the platforms – and the reviewing courts – in advance. With regard to the protection of fundamental rights, a major weakness of the DSA lies in the fact that the EU legislature arguably missed out on providing substantive guidelines in this regard (see IV.).

II. A Pluralist System of Fundamental Rights in Europe

1. The Plurality of Fundamental Rights Levels

a) Overlapping Spheres

One of the most distinctive characteristics of Europe's fundamental rights system is its pluralist structure. It is a system composed of different levels (or orders), encompassing national and regional fundamental rights as well as the EU Charter of Fundamental Rights (CFR) and also the European Convention on Human Rights (ECHR) as a pan-European minimum standard.[2] The particular challenge in handling this multi-level

[2] For further reading see the contributions in Matz-Lück and Hong (eds.), Grundrechte und Grundfreiheiten im Mehrebenensystem – Konkurrenzen und Interferenzen, Springer 2011 and Kleinlein, Grundrechtsföderalismus, Mohr Siebeck 2020.

system lies not so much in the plurality of levels as such, but in the fact that their scopes of application overlap to a considerable extent and these overlaps have to be federally coordinated and managed. It is here that mere plurality turns into (normative) pluralism.

While the German Federal Constitutional Court (FCC), to take one prominent example, had long advocated a rather strict separation of the spheres of EU and national fundamental rights, before eventually abandoning this approach in its landmark decisions on the *Right to be forgotten* in 2019,[3] The Court of Justice of the European Union (CJEU) had considered rather early that the spheres of EU and national fundamental rights overlap. The CJEU not only interpreted the scope of application of the Charter under Article 51 (1) CFR in a broad manner, holding that Member States are bound by the Charter to the extent that their actions fall within the scope of EU law.[4] The CJEU also allowed Member States to simultaneously apply national fundamental rights, provided that neither the level of protection under the Charter nor the primacy, unity, and effectiveness of EU law are thereby compromised (so-called *Melloni* conditions).[5] A parallel application of national fundamental rights presupposes that EU law does not fully determine or harmonize the case at hand, but leaves a margin of discretion to the Member States.[6] Recognising such parallel application essentially implies a change of perspective from separation to interaction, from dual to cooperative federalism.[7]

[3] German FCC, order of 6. November 2019, Case 1 BvR 16/13, *Right to be forgotten I* and order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*.

[4] CJEU, Case C-617/10, *Åkerberg Fransson*, EU:C:2013:105, para 21. Under Article 51 (1) CFR, Member States are only bound by the CFR when they are 'implementing' EU law. According to the CJEU in *Åkerberg Fransson*, situations in which Member States 'implement' EU law in the sense of Article 51 (1) CFR are situations falling within the scope of EU law. Hence, the 'applicability of European Union law entails applicability of the fundamental rights guaranteed by the Charter'. For a discussion see Reestman and Besselink, 'After *Åkerberg Fransson* and *Melloni*', 9 European Constitutional Law Review, 2013, 169; Sarmiento, 'Who's afraid of the Charter? The Court of Justice, national courts and the new framework of fundamental rights protection in Europe', 50 Common Market Law Review, 2013, 1267.

[5] CJEU, Case C-399/11, *Melloni*, EU:C:2013:107, para 60.

[6] See already CJEU, Case C-617/10, *Åkerberg Fransson*, EU:C:2013:105, para 29 ('not entirely determined') and, later, CJEU, Case C-476/17, *Pelham et al.,* EU:C:2019:624, para 81, according to which a parallel application of national fundamental rights 'is conceivable only in so far as those provisions do not effect full harmonisation'.

[7] In reference to Schütze, From Dual to Cooperative Federalism, OUP, 2009.

b) The Relevance of the Degree of Harmonisation of EU Law

Against this background, the degree of harmonisation of EU law and hence the reach of decision-making by the EU legislature play an important role in the EU's fundamental rights architecture. The degree to which the case at hand is determined by EU law determines the degree to which national fundamental rights may still apply: outside the scope of application of EU law, national fundamental rights continue to be the sole standard of review whereas, within the scope of application of EU law, they may apply alongside EU fundamental rights to the extent that EU law does not effect full harmonisation and the *Melloni* conditions are met.

Determining whether or not EU law effects full harmonisation or leaves a margin of discretion to the Member States is a difficult task. An appropriate test of discretion should, firstly, take account of the regulatory diversity of EU secondary law, secondly, be practically manageable and, thirdly, be compatible with different legal traditions in Europe.[8] Conceptually, such a test should focus on how broadly or deeply the EU legislature has exercised its policy-making power with regard to the Member States. The crucial question is therefore whether and to what extent the EU legislature has left it up to the Member States to take national measures under their own political responsibility within the framework of EU law.[9] This test does not replace, but rather presupposes the difficult and nitty-gritty work of interpreting each and every provision of EU secondary law. Also in the case law of the CJEU, the existence of a national margin of appreciation is increasingly founded on the intention of the EU legislature. Accordingly, the Court of Justice held in the copyright decisions *Funke Medien*, *Spiegel-Online,* and *Pelham* that the relevant provisions of EU legislation disclose "the intention of the EU legislature to grant a degree of discretion to the Member States".[10] A key in the Court's reasoning, therefore, is the degree of harmonisation as "intended by the EU legislature".[11] Following the reasoning of the CJEU in *Hernandez*, one might add that a margin of discretion exists where the EU legislature has decided to "grant the

[8] See on that Wendel, 'Europäischer Grundrechtsschutz und nationale Spielräume: Grundlagen und Grundzüge eines Spielraumtests im europäischen Grundrechtspluralismus', 57 Europarecht, 2022, 327, 353 et seq.

[9] Ibid. with more details.

[10] CJEU, Case C-469/17 *Funke Medien*, ECLI:EU:C:2019:623, para 34; Case C-516/17 *Spiegel-Online*, ECLI:EU:C:2019:625, para 23; Case C-476/17 *Pelham*, ECLI:EU:C:2019:624, para 82, On the decisions see Garben, 'Fundamental rights in EU copyright harmonization: Balancing without a solid framework', 57 Common Market Law Review, 2020, 1909.

[11] CJEU, Case C-469/17 *Funke Medien*, ECLI:EU:C:2019:623, para 40; Case C-516/17 *Spiegel-Online*, ECLI:EU:C:2019:625, para 25.

Member States an *option of legislating by virtue of EU law*".[12] Also the German FCC convincingly focuses on the intention of the EU legislature as manifested in EU secondary law.[13]

As a result, determining the degree of harmonisation as intended by the EU legislature requires a *norm-by-norm* micro-analysis.[14] The practical relevance of the test of discretion is significant. According to the recent case law of the German FCC in *Right to be Forgotten I and II* it determines whether the FCC applies EU fundamental rights – one of the most important innovations in the FCC's history – or continues to apply predominantly national fundamental rights. *Right to be Forgotten I* was about the conditions under which a previously lawful publication of a newspaper crime report, mentioning a sentenced criminal by name, became unlawful in view of the time which had elapsed since the original publication. The case fell within the scope of the so-called "media privilege" under the relevant EU provisions on data protection which leave a margin of discretion to the Member States.[15] Hence, the FCC applied the fundamental rights of the German Basic Law.[16] In contrast the case of *Right to be Forgotten II* did not fall under the "media privilege" because it involved a search engine and not a content provider. Holding the view that the relevant provisions of EU data protection law effected full harmonisation, the FCC applied EU fundamental rights.[17]

Before the German FCC delivered its decisions in *Right to be Forgotten I and II*, several constitutional courts of other EU Member States, in particular Austria, Belgium, France, and Italy, had already accepted EU fundamental rights as a standard of constitutional review.[18] Having examined the characteristics and modalities of this important transnational development elsewhere in more detail,[19] it suffices to say at this point that even though these courts apply EU fundamental rights by rather different means and to different extents, a common pattern stands out. Constitutional courts

[12] CJEU, Case C-198/13 *Hernandez*, ECLI:EU:C:2014:2055, para 44 (emphasis added). In the specific case, the CJEU found that the relevant provision of EU law did not confer a margin of discretion to the Member States within the scope of application of EU law, but rather referred to areas outside of the latter – areas in which the Member States continued to have comprehensive regulatory competences.

[13] German FCC, order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*, para 79.

[14] Cf. Conclusions of GA Bobek in Case C-310/16 *Dzivev*, ECLI:EU:C:2018:623, paras 73-74 and similarly German FCC, order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*, para 78.

[15] German FCC, order of 6. November 2019, Case 1 BvR 16/13, *Right to be forgotten I*, para 74 already with regard to Article 85 of the General Data Protection Regulation (GDPR).

[16] Ibid, paras 75-154.

[17] German FCC, order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*, paras 95 et seq.

[18] German FCC, case 1 BvR 276/17, *Right to be forgotten II*, 2019, para 50.

[19] Wendel, 'The Two-Faced Guardian: or How One Half of the German Federal Constitutional Court Became a European Fundamental Rights Court', 57 Common Market Law Review, 2020, 1383-1426.

have integrated EU fundamental rights – and often also the ECHR – into the standard of constitutional review, thereby simultaneously readjusting their relationship with the ordinary courts and (re-)establishing themselves as the central authorities of fundamental rights review at the national level.[20] The question of whether or not EU law effects full harmonisation plays a prominent role in this case law, albeit in different ways.

2. The Plurality of Standards

As a result of the different but overlapping levels of fundamental rights protection, there is also a plurality of potentially applicable standards. The standards specifically relevant to content regulation will be discussed in the second part of this paper. For now, it suffices to state in more general terms that differences in the scope of protection can be levelled to a certain extent by the principle of consistent interpretation. The Charter provides for an interpretation of EU fundamental rights in accordance with the ECHR (Article 52 (3) CFR) and in harmony with the constitutional traditions of the Member States (Article 52 (4) CFR). If a national court establishes that the potentially applicable fundamental rights standards coincide on the substance, it may even refrain from determining which fundamental rights regime – European or national – would actually have to be applied. As the German FCC's ruling in *Ecotoxicity Data* demonstrates, a national court may apply both standards simultaneously in such a case without having to decide whether the conditions for the application of one or the other standard are actually met.[21] To the extent that the norms converge as to their substance, the test of discretion becomes obsolete in legal practice. However, despite such tools for convergence, in many cases there are still considerable differences between the fundamental rights standards of the different levels. Such differences usually require a precise determination of the applicable standard in each case.

A pivotal key to the protection of fundamental rights at all levels is the principle of proportionality which is a "common European principle".[22] Despite all doctrinal and

[20] Ibid.

[21] German FCC, order of 27. April 2021, Case 2 BvR 206/14, *Ecotoxicity Data*.

[22] See Sommermann, 'Prinzipien des Verwaltungsrechts', in v. Bogdandy et al. (eds), Ius Publicum Europae Vol. V, CF Müller 2014, § 86, para 35. However, the fact that it is a common principle does not mean that it would be located at a meta-level. Rather, the principle is normatively anchored at different levels (European and national) at the same time.

subtle differences from country to country and from level to level,[23] the process of balancing, which is at the core of the proportionality test, is a crucial mechanism for adjusting and weighing the conflicting legal principles and interests.

3. Plurality of Relations, Horizontal Effect, and Legislative Responsibility

a) Starting Point

The system of fundamental rights protection in Europe is characterised not only by a plurality of levels and standards but also by a plurality of fundamental rights relations. The basic scenarios in which fundamental rights are applied – vertical, horizontal, multipolar, etc. – are familiar from the domestic legal framework. In a multi-level system, however, these fundamental rights relationships grow more complex. This is even true for the classic vertical relationship between the citizens and public authority. In the case of multi-level legislation, which runs from the EU legislature to the national legislature and, in some cases, also to the regional legislature, this relationship is ultimately layered. It is regularly necessary to determine precisely the exercise of which public authority – European, national, or both – is being challenged by an individual on the basis of his or her fundamental rights. Or to frame it differently, there is a plurality of vertical relations.

b) Horizontal Effect of EU Fundamental Rights: Basic Conditions

Even more complicated are horizontal or multipolar fundamental rights relations, emanating from disputes between two private parties or from disputes involving three (private or public) parties or more. In such cases it is precisely the balancing of conflicting fundamental rights of different levels that can raise particular difficulties, as famously shown by the interplay between the German FCC and the ECtHR in the *Caroline von Monaco* cases.[24] Regarding the horizontal effect of *EU* fundamental rights in particular, crucial questions are still unresolved. The conditions and the nature of the horizontal effect of EU fundamental rights have at best been rudimentarily clarified by the case

[23] For instance, the structure of the proportionality test under Article 52 (1) CFR slightly differs from that under the Basic Law.

[24] ECtHR (GC), No. 59320/00 – *von Hannover/Germany I* and No. 40660/08 – *von Hannover/Germany II* and on the other side German FCC, BVerfGE 101, 361 *Caroline of Monaco I* and BVerfGE 120, 180 *Caroline von Monaco II*.

law and remain the subject of ongoing discussions.[25] What can be inferred with a sufficient degree of certainty from the existing case law is that some EU fundamental rights can indeed have horizontal effect, particularly in the areas of labour and non-discrimination law.[26] The CJEU convincingly opposed an interpretation, according to which Article 51 (1) FCR would generally exclude such horizontal effect. According to the Court of Justice,

"although Article 51(1) of the Charter states that the provisions thereof are addressed to the institutions, bodies, offices, and agencies of the European Union ... and to the Member States only when they are implementing EU law, Article 51(1) does not, however, address the question of whether those individuals may, where appropriate, be directly required to comply with certain provisions of the Charter and cannot, accordingly, be interpreted as meaning that it would systematically preclude such a possibility."[27]

Hence, Article 51 (1) CFR does not provide an answer to the question of horizontal effect and arguably was not intended to do so. It addresses primarily the vertical relationship between the EU and the Member States[28] and allocates the responsibility for respecting fundamental rights to the respective level of governance.[29]

It follows from the CJEU's case law that the horizontal effect of a fundamental right enshrined in the Charter presupposes that this right "is sufficient in itself to confer on

[25] See on this topic Frantziou, 'The horizontal effect of fundamental rights in the European Union: a constitutional analysis', OUP 2019; id., 'The Horizontal Effect of the Charter: Towards an Understanding of Horizontality as a Structural Constitutional Principle?', 22 Cambridge Yearbook of European Legal Studies, 2020, 208; S Prechal, 'Horizontal direct effect of the Charter of Fundamental Rights of the EU', 66 Revista de Derecho Comunitario Europeo, 2020, 407; Jarass, 'Die Bedeutung der Unionsgrundrechte unter Privaten', Zeitschrift für Europäisches Privatrecht, 2017, 311, 330 et seq.; Seifert, 'L'effet horizontal des droits fondamentaux – Quelques réflexions de droit européen et de droit comparé', Revue trimestrielle de droit européen, 2012, 801; Safjan, 'The Horizontal Effect of Fundamental Rights in Private Law – On Actors, Vectors, and Factors of Influence', in Liber Amicorum for Hans Micklitz, Springer 2014, 123; C Unseld, Zur Bedeutung der Horizontalwirkung von EU-Grundrechten, Mohr Siebeck 2018.

[26] CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257 (with regard to Articles 21 and 47 CFR); Case C-68/17 *IR Catholic Chief physician* ECLI:EU:C:2018:696 (with regard to Article 21 CFR); Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871 (with regard to the right to paid annual leave under Article 31 (2) CFR); Case C-684/16 *Max-Planck* ECLI:EU:C:2018:874 (also with regard to Article 31 (2) CFR). For the case law before the entry into force of the Charter see, in particular, CJEU, Case C-144/04 *Mangold* ECLI:EU:C:2005:709, para 77.

[27] CJEU, Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871, para 87.

[28] Conclusions of AG Cruz Villalón in case C-176/12 *Association de mediation sociale*, paras 28-32.

[29] Lenaerts, 'The Role of the Charter in the Member States', in Bobek and Adams-Prassl (eds.), The EU Charter of Fundamental Rights in the Member States, Bloomsbury 2020, 19, 25 .

individuals a right which they may rely on as such in a dispute with another individual".[30] By contrast, a right that must be given more specific expression in EU or national law cannot be invoked in a dispute between individuals, as is the case with the workers' right to information and consultation under Article 27 CFR.[31] Hence, the conditions of horizontal effect correspond partly, but not entirely, to the requirements of direct effect.[32] The special feature of fundamental rights with horizontal effect seems to be that they are mandatory individual rights which, by their "very nature", entail "a corresponding obligation" on the other private party, as is the case with the right to paid annual leave under Article 31 (2) CFR.[33] Hence, what is decisive is the substantive nature of the respective fundamental right.[34]

In order to justify the horizontal effect of non-discrimination rights, the CJEU also draws on its classic jurisprudence on the prohibitions of discrimination under Article 18 TFEU, Article 157 TFEU, and the fundamental freedoms,[35] arguing that the mere fact that such provisions are addressed principally to the Member States does not preclude the application to relations between individuals.[36] As regards the prohibition of discrimination under Article 21 CFR, the CJEU holds that this provision is, as regards its mandatory effect, "no different, in principle, from the various provisions of the founding Treaties prohibiting discrimination on various grounds, even where the discrimination derives from contracts between individuals".[37].

c) Distinguishing Direct from Indirect Horizontal Effect: A Quarrel About Nothing?

Establishing that EU fundamental rights can have a horizontal effect does not answer the question of whether this effect is *direct* or *indirect*. There are several substantial arguments against a direct horizontal effect of fundamental rights, including that such effect could lead to undue limitations of private autonomy and liberties,[38] that direct

---

[30] Ibid, para 89 and CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 76.

[31] CJEU, Case C-176/12 *Association de médiation sociale* ECLI:EU:C:2014:2, para 47.

[32] Similarly Prechal, 'Horizontal direct effect of the Charter of Fundamental Rights of the EU', 66 Revista de Derecho Comunitario Europeo, 2020, 407, 420-421.

[33] CJEU, Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871, para 90.

[34] Prechal, 'Horizontal direct effect of the Charter of Fundamental Rights of the EU', 66 Revista de Derecho Comunitario Europeo, 2020, 407, 419.

[35] Pars pro toto CJEU, case 43/75 *Defrenne II* ECLI:EU:C:1976:56, para 39; Case C-281/98 *Angonese* ECLI:EU:C:2000:296, paras 33-36; case C-411/98 *Ferlini*, ECLI:EU:C:2000:530, para 50; Case C-438/05, *Viking* ECLI:EU:C:2007:772, paras 57-61.

[36] CJEU, Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871, para 88.

[37] CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 77.

[38] See with specific regard to digital intermediary services Wielsch, 'Verantwortung von digitalen Intermediären für Rechtsverletzungen Dritter' 10 Zeitschrift für Geistiges Eigentum, 2018, 1, 34.

horizontal effect is not common to most constitutional systems in Europe[39] and that private parties could not themselves satisfy the requirement of Article 52 (1) CFR, according to which any limitation on the exercise of EU fundamental rights must be "provided for by law".[40] In order to address these objections while at the same time taking into account the fundamental normative impact of fundamental rights on the legal order as a whole and the fact that individual freedom and equality can also be threatened by private actors, especially if they are in a position of structural superiority (as is particularly the case with platforms in the digital space), many scholars make the case for an indirect horizontal effect.[41] In the case law of the German FCC this approach is known since the late 1950s as the so-called *mittelbare Drittwirkung*.[42]

However, in the CJEU's case law, there is no explicit doctrine of either direct or indirect horizontal effect. EU fundamental rights can impact legal disputes between private parties in rather different ways, varying from case to case. The most common *modus operandi* is that the relevant EU legislation – and the national law implementing it – have to be interpreted in conformity with EU fundamental rights. This may require the national authorities to interpret EU law and the national provisions transposing it in a way "which allows a fair balance to be struck between the various fundamental rights protected" by EU law.[43] This approach, which focuses on the consistent interpretation of civil law provisions with EU fundamental rights, essentially corresponds to an indirect horizontal effect.[44]

In other cases, the CJEU has allowed private parties to rely directly on EU fundamental rights vis-à-vis other private parties. However, the effects of such a direct invocation of the Charter also vary.[45] In a first and rather typical scenario, the application of a Charter right entails the non-application of conflicting national law, supposing that

[39] For a comparative overview see Besselink, 'The Protection of Fundamental Rights post-Lisbon', in Laffranque (ed.), Reports of the FIDE Congress Tallinn 2012 – Vol. 1, Tartu University Press 2012, 18-19.

[40] Conclusions of AG Trstenjak in case C-282/10 *Dominguez*, ECLI:EU:C:2011:559, para 83.

[41] See, pars pro toto, Jarass, 'Die Bedeutung der Unionsgrundrechte unter Privaten', Zeitschrift für Europäisches Privatrecht, 2017, 311, 331 with further references. With specific regard to platform regulation see Denga, 'Plattformregulierung durch europäische Werte: Zur Bindung von Meinungsplattformen an EU-Grundrechte', 56 Europarecht, 2021, 569, 595.

[42] German FCC, judgment of 15. January 1958, Case 1 BvR 400/51 *Lüth* paras 25 et seq.

[43] CJEU, Case C-275/06 *Promusicae* ECLI:EU:C:2008:54, para 70.

[44] Cf. German FCC, order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*, para 9, concluding that EU fundamental rights ultimately have a horizontal effect comparable to that of German fundamental rights.

[45] Cf. Prechal, 'Horizontal direct effect of the Charter of Fundamental Rights of the EU', 66 *Revista de Derecho Comunitario Europeo*, 2020, 407, 410.

the latter cannot be interpreted in conformity with EU fundamental rights.[46] In a second scenario, private party A invokes a Charter right which becomes, in turn, a direct source of an obligation of private party B. Both scenarios can also be combined.[47] Both scenarios demonstrate that drawing a sufficiently clear line between direct and indirect horizontal effects is a difficult, perhaps even unnecessary task. The "negative" effect produced in the first scenario must not necessarily be understood as direct, but can also be conceived as an indirect one, as it only leads to the non-application of national legislation. In contrast, the "positive" effect in the second scenario seems to be the very prototype of direct horizontal effect, since the fundamental right invoked by private party A automatically results in an obligation for private party B. However, if one assumes that it is ultimately up to the courts to establish such an obligation – for instance that a platform must delete content uploaded by a user – as the result of an exercise of balancing, even such a "positive" effect could be considered indirect. For the individuals concerned, it makes little difference whether they lose the case because of an obligation that is directly incumbent on them on the basis of a fundamental right or because of the non-application of national legislation that would be favourable to them. As a *modus operandi*, the negative effect is by no means less powerful or intrusive than the positive one, a fact that is also reflected in the fierce (judicial) opposition to some CJEU rulings at the national level.[48] Ultimately, negative and positive effect appear to be two sides of the same coin. Or to frame it differently, the differences appear more gradual, not principled. Even in the recent jurisprudence of the German FCC, i.e. the court which has invented *mittelbare Drittwirkung*, the boundary between direct and indirect horizontal effect has become more and more blurred.[49]

Overall, it seems that the exact classification of the horizontal effect – direct or indirect – is less important than first appearances might suggest.[50] It might be too much to

[46] See, as an example, CJEU, Case C-684/16 *Max-Planck* ECLI:EU:C:2018:874 and CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 82.

[47] As was the case in CJEU, Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871, para 92. Here, the (negative) obligation of the national courts to disapply the conflicting national legislation was combined with the (positive) obligation to ensure that the legal heir receives payment from the employer.

[48] See the *ultra-vires*-decision of the Danish *Højesteret*, case 15/2014 *Ajos*, 2016, with regard to the CJEU's *Mangold* ruling and, still pending, German FCC, Case 2 BvR 934/19 *Egenberger*.

[49] See German FCC BVerfG, order of 11. April 2018, Case 1 BvR 3080/09 *Football Stadium Ban* and the analysis of Kulick, Horizontalwirkung im Vergleich, Mohr Siebeck, 207-214. The FCC itself admits that the indirect horizontal effect 'can ultimately be close, or even equal to, its binding effect on the state', see German FCC, order of 6. November 2019, Case 1 BvR 16/13, *Right to be forgotten I*, para 88.

[50] Similarly Jarass, 'Die Bedeutung der Unionsgrundrechte unter Privaten', Zeitschrift für Europäisches Privatrecht, 2017, 311, 332-333.

say that this is a quarrel about nothing, a *querelle d'allemand*. However, in terms of constitutional law, the key problem is a matter of competence.

d) A Matter of Competence: Legislative Responsibility for Fundamental Rights

From a constitutional point of view, the central question is less one of the proper classification of horizontal effects than one of competence, namely the question of how far it is up to the EU legislature to already strike a balance between the conflicting fundamental rights. To what extent is it incumbent on the European legislature to provide a substantive legislative framework for the balancing of the conflicting fundamental rights? It is a major question of power-sharing between the legislature on the one hand and other actors, such as private platforms and the reviewing courts on the other. The less the legislature outlines or predetermines the horizontal or multipolar fundamental rights relationship, the more leeway is ultimately left to other actors and ultimately the courts in weighing up the conflicting fundamental rights. Against this background, the CJEU has not only stated that the (judicial) obligation to ultimately strike a balance between conflicting fundamental rights does by no means prevent the possibility of individuals invoking these fundamental rights in horizontal disputes.[51] It has also made clear that the national courts must take into consideration the balance struck between those interests by the EU legislature.[52]

But to what extent is the EU legislature under a constitutional obligation to strike this balance itself? Here, too, many things are still unresolved. And although already offering some guidance, the existing case law – to which we will return in the context of the DSA – does not allow a straightforward conclusion as to the extent to which the CJEU holds the Union legislature responsible to regulate the essential questions of the exercise of fundamental rights in horizontal (or multipolar) fundamental rights.

4. Plurality in Enforcement

Beyond the plurality of levels, standards, and relations, the European system of fundamental rights in the EU is also characterised by a plurality in enforcement. Just as the EU generally does not follow the model of dual but of cooperative federalism,[53] its legal system is also a mixed one. In this judicial system it is for both the CJEU and the national courts to ensure the full application of EU law as well as the effective judicial

[51] CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 80.

[52] Ibid, para 81.

[53] See again Schütze, From Dual to Cooperative Federalism, Oxford 2009.

protection of fundamental rights.[54] Within this mixed and cooperative framework, national courts play a pivotal role in enforcing EU law across the Union. They have a *European mandate*[55] in the sense that they fulfil the "role of a court of the European Union", as the CJEU has recently put it.[56] This European mandate covers the interpretation of EU law – in dialogue with the Court of Justice[57] – as well as the effective application of EU law, including the non-application of conflicting national law, even if the latter enjoys constitutional rank.[58]

The national courts' European mandate is crucial to the functioning of the European community of law, as demonstrated by the fact that EU law requires national courts to meet certain minimum standards of judicial independence under Article 19 (1) TEU, particularly in the context of the rule of law crisis.[59] However, the rule of law crisis in Poland and other countries also shows how diverse and vulnerable the decentralised system of enforcement turns out to be. But apart from the systemic dismantling of the rule of law in some member states, the judicial protection of fundamental rights is also organised very differently in the EU. This is true for the existence or non-existence of specialised constitutional courts as well as for the role of the judiciary vis-à-vis the other branches of power in the respective constitutional system.

[54] Cf. CJEU, Case C-357/19 et al. *Euro Box Promotion et al.* ECLI:EU:C:2021:1034, para. 254 with further references.

[55] In detail Claes, 'The National Courts' Mandate in the European Constitution', Oxford 2005, 58 et seq.; Temple Lang, 'The Duties of National Courts under the Constitutional Law of the European Community', Exeter 1987.

[56] CJEU, Case C-357/19 et al. *Euro Box Promotion et al.* ECLI:EU:C:2021:1034, para 257.

[57] While the interpretation of EU law is initially a matter for the national courts, the Court of Justice has exclusive jurisdiction to give the definitive and finally binding interpretation of EU law, see ibid, para. 254.

[58] CJEU, case 106/77 *Simmenthal II*, 1978, para. 24; Case C-573/17 *Popławski,* 2019, para. 61 et seq.; Case C-357/19 et al. *Euro Box Promotion et al.* ECLI:EU:C:2021:1034, paras 252, 254, 257 and Case C-430/21 *RS – Effect of judgments of a constitutional court* ECLI:EU:C:2022:99, para 53.

[59] See, pars pro toto, CJEU, Case C-64/16 *Associação Sindical dos Juízes Portugueses*, 2018; Case C-585/18 et al. *A.K.*, 2019; Case C-791/19 *Commission/Poland – Disciplinary regime applicable to judges*, 2021.

III. Fundamental Rights and Content Regulation under the DSA

1. Levels and Margins: How Much Scope for National Fundamental Rights?

Having explored the pluralist structure of the system of fundamental rights in Europe, we will now turn towards a more specific analysis of fundamental rights protection with regard to the DSA. A primary issue concerns identifying the levels of fundamental rights protection – national, European, and international – that are relevant in the context of the DSA. It should first be noted that EU fundamental rights must be observed throughout. Due to Article 51 (1) CFR, the EU legislature is already bound by EU fundamental rights in designing and adopting the DSA. The Commission is also bound by EU fundamental rights, which can become relevant in many respects in the context of the DSA. Furthermore, within the scope of application of the DSA, also the EU Member States are bound by EU fundamental rights since they are "implementing" Union law in the sense of Article 51 (1) CFR. The question of whether and to what extent private parties – and platforms in particular – are bound by EU fundamental rights is something we will return to in detail in the context of fundamental rights relations.

The ECHR is also of utmost importance, because the interpretation of key EU fundamental rights, such as freedom of expression and information (Article 11 CFR) or the right to privacy (Article 7 CFR), is guided by the corresponding ECHR rights, pursuant to Article 52 (3) CFR. This, of course, does not prevent EU law from providing more extensive protection.[60] Recital 47 of the DSA also mentions that all providers of intermediary services shall "pay due regard to relevant international standards for the protection of human rights, such as the United Nations Guiding Principles on Business and Human Rights." Of course, this is merely a declaratory reference.

More difficult to answer is the question of how far *national fundamental* rights do apply alongside EU fundamental rights within the framework of the DSA. As demonstrated above,[61] the degree to which the case at hand is determined by EU law determines the degree to which national fundamental rights may still apply. National fundamental rights apply alongside EU fundamental rights, provided that EU (secondary) law does not effect full harmonisation and the *Melloni* criteria are met.[62] Determining

[60] See explicitly Article 52 (3) sentence 2 CFR.

[61] See 2.1.2.

[62] Ibid.

the degree of harmonisation as "intended by the EU legislature" [63] always requires, as we have seen, a *norm-by-norm* analysis. The core of this analysis is to ask whether and to what extent the EU legislature has left it up to the Member States to take national measures under their own political responsibility within the framework of EU law. The type of EU legislation as such does not allow for definitive conclusions, as directives can set definitive standards, just as regulations can contain opening clauses or other mechanisms that provide latitude at the national level.[64]

Hence, the mere fact that the DSA is a regulation in the sense of Article 288 (2) TFEU does not preclude national margins of discretion. Certainly, the DSA sets a trend towards centralization. According to its Recital 9, which received its current wording rather late in the legislative process, the DSA "fully harmonises the rules applicable to intermediary services". Consequently, "Member States should not adopt or maintain additional national requirements relating to the matters, unless explicitly provided for" in the DSA. Indeed, the latter will make existing national law obsolete to a considerable extent.[65] However, it is also clear that the harmonising effect of the DSA only extends to its scope of application and does not, even within this scope, apply without exception ("unless explicitly provided for").

A particularly far-reaching exception relates to the executive enforcement of the DSA at the national level. Under Article 51 (6) DSA it is up to the Member States to lay down "specific rules and procedures for the exercise" of the executive powers of the independent national bodies (so-called services coordinators) and to "ensure that any exercise of those powers is subject to adequate safeguards laid down in the applicable national law in compliance with the Charter and with the general principles of Union law". And under Article 52 DSA it is, again, up to the Member States to "lay down the rules on penalties applicable to infringements of this Regulation by providers of intermediary services". The DSA thus largely leaves it to the national legislatures to lay out the normative frameworks of executive enforcement. In enacting these national legal frameworks, the national legislatures are bound by the Charter *and* national fundamental rights, provided that the *Melloni* conditions are met.[66] Another important area in which national fundamental rights continue to apply in part are the rules that determine the legality of content. This question falls, in principle, outside the scope of the

[63] CJEU, Case C-469/17 *Funke Medien*, ECLI:EU:C:2019:623, para 40; Case C-516/17 *Spiegel-Online*, ECLI:EU:C:2019:625, para 25.

[64] German FCC, order of 6. November 2019, Case 1 BvR 276/17, *Right to be forgotten II*, para 79.

[65] See the contributions on France, Germany and Italy in this volume. For Germany see also Janal, 'Friendly Fire? Das Urheberrechts-Diensteanbieter-Gesetz und sein Verhältnis zum künftigen Digital Services Act', Gewerblicher Rechtsschutz und Urheberrecht, 2022, 211.

[66] See, again, 2.1.2.

DSA. As far as the national legislation determines which content is illegal,[67] this legislation must be in conformity with national fundamental rights.

2. Relevant Fundamental Rights Standards in the Context of the DSA

The fundamental rights standards that can become relevant within the scope of the DSA are manifold and vary from stakeholder to stakeholder.[68] Users can be affected particularly in their right to privacy and family life (Article 7 CFR), their right to the protection of personal data (Article 8 CFR), their freedom of expression and information (Article 11 CFR), their right to non-discrimination (Article 21 CFR) and their right to a fair trial and effective remedy (Article 47 CFR).

Service providers can be affected in their freedom of occupation (Article 15 CFR); their freedom to conduct a business, including the freedom of contract (Article 16 CFR); their freedom of property, including copyright (Article 17 CFR); and also their right to a fair trial and effective remedy (Article 47 CFR). Furthermore, they act as gatekeepers for mass communication by filtering and listing content and make use, even if it happens through algorithms, of their freedom of media and also of their freedom of expression (Article 11 CFR).[69]

Other stakeholders, and particularly private parties affected by illegal or "lawful but awful" content can be affected in human dignity (Article 1 or 4 CFR); their right to privacy (Article 7 CFR);[70] their freedom of information (Article 11 CFR); their right to protection of property, including intellectual property (Article 17 CFR);[71] their right to non-discrimination (Article 21 CFR); the rights of the child (Article 24 CFR); and, like the others, their right to a fair trial and effective remedy (Article 47 CFR).

Most of these rights have to be applied in harmony with the ECHR.[72] With the notable exception of Articles 1 and 4 CFR, which enjoy absolute protection, all of the rights can be weighed against each other in different scenarios. This can lead to quite different results from case to case. It follows from the preceding case law, for instance,

---

[67] Cf. Article 3 lit. h) DSA: '... not in compliance with Union law or the law of any Member State which is in compliance with Union law' (emphasis added).

[68] See also recital 52 of the DSA.

[69] Schiedermair ad Weil, 'Online-Intermediäre als Träger der Meinungsfreiheit', 75 Die Öffentliche Verwaltung, 2022, 305, 307 et seq.; Paal, 'Vielfaltssicherung bei Intermediären', Multimedia und Recht, 2018, 567, 568 with regard to national law.

[70] See Gersdorf, 'Artikel 7 GRCh', in: id and Paal (eds.), BeckOK Informations- und Medienrecht, 2021, para 23.

[71] See Lock, 'Article 17 CFR', in: Kellerbauer et al. (eds.), The EU Treaties and the Charter of Fundamental Rights, 2019, para 11.

[72] See, pars pro toto, CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, paras 44-46, 68 with regard to Article 11 CFR.

that (*de facto*) obliging service providers to use upload filters for the purpose of copyright can be proportionate under certain conditions with regard to Article 11, Article 16, and Art. 17 CFR.[73] Furthermore, a general filtering requirement for service providers for the purpose of copyright protection[74] would be incompatible with Article 8, Article 11 and Article 16 CFR.[75] The prohibition to impose a general obligation of monitoring or active searching is also contained in Article 8 DSA and, until 16 February 2024, in Article 15 of the E-Commerce-Directive.[76]

3. Fundamental Rights Relations

a) Does the DSA Establish a Direct Horizontal Effect of Fundamental Rights?

At first glance, it looks as if the DSA imposes far-reaching fundamental rights obligations on platforms. One of the most controversial issues of the DSA is whether or not it establishes a direct horizontal effect of EU fundamental rights.[77] At the centre of this discussion is Article 14 (4) DSA, formerly Article 12 (2) of the initial proposal.[78] This provision is part of the Article on the terms and conditions of providers. It requires providers to "act in a diligent, objective and proportionate manner in applying and enforcing the restrictions" on the use of their service, such as removing or demoting content,

"with due regard to the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service, such as the freedom of expression, freedom and pluralism of the media, and other fundamental rights and freedoms as enshrined in the Charter."

---

[73] CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, paras 82-97.

[74] I.e. a monitoring obligation that is non-specific, relates to any future infringement, is unlimited in time and also covers works that have not yet been created.

[75] CJEU, Case C-70/10 *Scarlet Extended* ECLI:EU:C:2011:771, paras 45-54.

[76] Directive 2000/31. See also CJEU, Case C-18/18 *Eva Glawischnig-Piesczek* ECLI:EU:C:2019:821.

[77] Eifert et al., 'Taming the Giants: the DMA/DSA Package', 58 Common Market Law Review, 2021, 987, 1012-1014; Denga, 'Plattformregulierung durch europäische Werte: Zur Bindung von Meinungsplattformen an EU-Grundrechte' 56 Europarecht, 2021, 569, 590-596; Appleman et al., 'Article 12 DSA: Will platforms be required to apply EU fundamental rights in content moderation decisions?', DSA Observatory of 31 May 2021, available at https://dsa-observatory.eu/2021/05/31/article-12-dsa-will-platforms-be-required-to-apply-eu-fundamental-rights-in-content-moderation-decisions/.

[78] COM 2020, 825 final. This is why several early comments or papers refer to Article 12 (2) DSA proposal.

Does Article 14 (4) DSA have "revolutionary potential" as it is sometimes said to do?[79] Does it provide a legal foundation for directly binding the platforms to EU fundamental rights and establishing a direct horizontal effect of EU fundamental rights? The better arguments suggest otherwise. It follows from the architecture of the European constitutional order that EU legislation cannot determine the effects of EU fundamental rights and in particular not the question of who is bound by the Charter. This question is exclusively determined by EU primary law. This is not to say that EU legislation does not have an impact on EU fundamental rights. On the contrary, it does to a considerable extent. The EU legislature may (and sometimes must) concretise EU fundamental rights in substantive terms.[80] For instance, Article 21 CFR is given concrete expression in Directive 2000/78.[81] Other fundamental rights, such as the right to property, require *per definitionem* a concretisation by means of legislation.[82] Furthermore, the extent to which the Member States must respect the Charter depends on the scope of EU legislation, given that Article 51 (1) CFR establishes that the Member States are only bound by the Charter when implementing EU law. As CJEU President Koen *Lenaerts* once put it, the Charter follows EU (secondary) law like a "shadow".[83] However, both forms of influence of secondary law on the precise content and scope of EU fundamental rights have to be separated from the question of who is bound by the Charter.

The question of who is bound by EU fundamental rights relates to the normative effects of a key element of EU constitutional law and is exclusively determined by EU primary law. Otherwise, the normativity of EU fundamental rights as elementary guarantees of freedom and equality would be called into question. If the EU legislature had the legal power to (positively) establish the binding force of the Charter rights through legislation, it would also have to have the power to (negatively) repeal it. Consequently, the EU legislature, which as a public authority is itself fully bound by EU fundamental rights under EU constitutional law, would ultimately be in a position to dispose of the scope of these core guarantees, including its own fundamental rights obligations. This would undermine the essential function of EU fundamental rights as inescapable elementary guarantees of individuals.

---

[79] Quintais et al., 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation', German Law Journal, 2023, forthcoming.

[80] See as an example CJEU, Case C-176/12 *AMS*, ECLI:EU:C:2014:2, paras 45 et seq.

[81] Already before the entry into force of the Charter in 2009, the right to non-discrimination was acknowledged as a general principle of EU law.

[82] In detail Michl, 'Unionsgrundrechte aus der Hand des Gesetzgebers', Mohr Siebeck, 2018.

[83] Lenaerts, Gutiérrez-Fons, 'The Place of the Charter in the EU Edifice', in Peers et al. (eds.), The EU Charter of Fundamental Rights: A Commentary, Hart 2014, 1560, 1568.

Against this background, it is not convincing on principle grounds of EU constitutional law to interpret Article 14 (4) DSA as a legal basis for establishing a direct horizontal effect of EU fundamental rights. Article 14 (4) DSA could, of course, trigger an indirect horizontal effect insofar as it establishes certain obligations of platforms – obligations which must be interpreted and put into practice in the light of EU fundamental rights.[84] Even for this purpose, however, Article 14 (4) DSA is kept remarkably general and vague, a point to which we will return.

It is important to note that the rejection of Article 14 (4) DSA as a legal basis for establishing direct horizontal effect in no way implies that the Charter rights in question could in fact have no (direct) horizontal effect. On the contrary, there are strong arguments why several, if not the majority of the above-mentioned EU fundamental rights could actually have (direct) horizontal effect. As explained above, the CJEU has already recognised the horizontal effects of Article 21 CFR.[85] And as we have seen, classifying the horizontal effects acknowledged in the existing case law is difficult, because the differences seem rather gradual than principled.[86] In particular, the rather typical *modus operandi*, that the (direct) reliance on a Charter right in litigation between private parties leads to the non-application of national law,[87] can, for good reasons, also be conceived as an indirect horizontal effect.[88]

Be that as it may, recognising such horizontal (direct) effect seems particularly plausible also for the freedoms under Articles 7, 8, and 11 CFR. These fundamental rights not only fulfil the basic criteria of the case law in that they are sufficient in themselves to confer on an individual a right which may, by its very nature, entail a corresponding obligation on another private party.[89] Moreover, these freedoms are exercised to a significant degree in virtual space, a space where service providers and in particular large platforms enjoy structural superiority over users. In order to keep the normative promise which is inherent in the guarantee of freedoms such as those enshrined in Articles 7, 8, and 11 CFR (and other fundamental rights), it seems appropriate, if not necessary, to include service providers, and in any case large platforms, in the group of those

[84] Similarly Denga, 'Plattformregulierung durch europäische Werte: Zur Bindung von Meinungsplattformen an EU-Grundrechte', 56 Europarecht, 2021, 569, 594.

[85] See 2.3.2. and particularly CJEU Case C-414/16 *Egenberger* ECLI:EU:C:2018:257.

[86] See 2.3.3.

[87] See again as an example CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 82 and Case C-684/16 *Max-Planck* ECLI:EU:C:2018:874.

[88] See 2.3.3.

[89] CJEU, Case C-569/16 et al. *Bauer and others* ECLI:EU:C:2018:871, para 89-90; CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 76.

bound by said EU fundamental rights.[90] Thus, apart from the reasons relating to the specific nature of the fundamental rights in question, there are also teleological reasons for recognizing a (direct) horizontal effect. However, all of these reasons are normatively founded in the constitutional system of EU fundamental rights and the nature of the respective fundamental rights themselves and *not in secondary law*. The fact that Article 14 (4) DSA refers to the compliance of providers with EU fundamental rights can, of course, ultimately be understood as a declaratory acknowledgement, on the part of the EU legislature, that the relevant EU fundamental rights have by themselves (direct) horizontal effect.[91] While the exact classification of the horizontal effect seems ultimately less important, the question of the extent to which the EU legislature has an obligation to reconcile conflicting fundamental rights in statutory law, i.e. the extent to which there is not only the possibility but an obligation to predetermine the balancing of conflicting fundamental rights, seems crucial in terms of constitutional law.

b) Taking or Escaping Legislative Responsibility for Fundamental Rights

Did the EU legislature take or escape its legislative responsibility for fundamental rights in the case of the DSA? To what extent was it under a constitutional obligation to predetermine the balance between the conflicting fundamental rights of platforms, users, and other stakeholders in advance and, particularly, to limit the leeway of platforms and ultimately the courts in advance? There is no case law on the DSA yet. And while many things are still unresolved in the existing case law, some important lines of reasoning can be drawn from it. For example, the CJEU has established minimum requirements as to the substance of legislation that interferes with fundamental rights. In its landmark ruling of 26 April 2022 on the new copyright liability of service providers under the DSM Directive,[92] which *de facto* imposes the use of upload filters on service providers, the CJEU consolidated his preceding case law, stating that EU legislation which

"entails an interference with fundamental rights must lay down *clear and precise rules governing the scope and application of the measure* in question and *imposing minimum safeguards* so that the persons whose exercise of those rights is limited have sufficient guarantees to protect them effectively against the risk of abuse. That legislation must, in particular, indicate *in what circumstances and under which conditions* such a

[90] Similarly Prechal, 'Horizontal direct effect of the Charter of Fundamental Rights of the EU', 66 Revista de Derecho Comunitario Europeo, 2020, 407, 418-419. Cf. also CJEU, Case C-131/12 *Google Spain* ECLI:EU:C:2014:317, paras 80-81.

[91] This is mirrored in the fact that the Commission seems to almost naturally regard providers as entities bound by EU fundamental rights in the impact assessment of the DSA COM, 2020, 825 final, p. 14-15.

[92] Directive 2019/790.

measure may be adopted, thereby ensuring that the interference is limited to what is strictly necessary. The need for such safeguards is *all the greater where the interference stems from an automated process*".[93]

If an EU legislative act interferes with fundamental rights but remains too vague and general, it may even be invalid, as was the case with the former Data Retention Directive.[94] This line of reasoning refers to the necessary extent to which interference with fundamental rights resulting from an EU legislative act must be specified and limited in that same act in order to be proportionate[95] or to fulfil the requirement "provided by law" under Article 52 (1) CFR.[96]

In the case that a balance must be struck between conflicting fundamental rights, the CJEU has explicitly recognised that the EU legislature may grant other players a considerable degree of flexibility in concretising such multipolar fundamental rights relations. This is not only true when it comes to protecting the *federal* diversity of fundamental rights, for instance when EU legislation leaves a margin of discretion to the national legislature and "does not itself effect the necessary reconciliation between [conflicting fundamental rights[97]], but merely provides a framework for the reconciliation which Member States must achieve between those two values."[98] It also applies to the obligations that the EU legislature imposes on service providers. With specific regard to obligations imposed on service providers in order to protect copyright regimes, the CJEU held that it may

"*even prove necessary* – in order to respect the freedom of those service providers to conduct a business, guaranteed in Article 16 of the Charter, and to respect the *fair balance* between that freedom, the right to freedom of expression and information of the

---

[93] CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, para 67 (emphasis added). For the preceding case law see, in particular, CJEU, Case C-311/18 *Schrems II [Privacy Shield]* ECLI:EU:C:2020:559, para 175-176.

[94] CJEU, Case C-293/12 et al. *Digital Rights Ireland Ltd and others* ECLI:EU:C:2014:238. The former Directive 2006/24/EC provided for a general obligation of telecommunications providers to retain connection data, without at the same time providing for sufficient procedural and substantive safeguards.

[95] This is the approach chosen by the CJEU, ibid, para 54 within the framework of the necessity test. See also CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, para 67.

[96] This is the – doctrinally more intriguing – approach of AG Cruz Villalón in Case C-293/12 et al. *Digital Rights Ireland Ltd and others*, ECLI:EU:C:2013:845, paras 117 et seq. that derives from Article 52 (1) CFR 'quality of law requirements', including minimum substantive requirements (ibid, paras 108-132) that are distinct from the proportionality test (ibid, paras 133-152).

[97] In this case animal welfare and the freedom to manifest religion.

[98] CJEU, Case C-336/19 *Centraal Israëlitisch Consistorie van België and Others* ECLI:EU:C:2020:1031, paras 47 and also 71.

users of their services, enshrined in Article 11 of the Charter, and the right to intellectual property of the rightsholders, protected in Article 17(2) of the Charter – *to leave those service providers to determine the specific measures to be taken* in order to achieve the result sought; accordingly, they can choose to put in place the measures which are best adapted to the resources and abilities available to them and which are compatible with the other obligations and challenges which they will encounter in the exercise of their activity."[99]

This precedent could indicate that the high degree of flexibility granted by the EU legislature to providers in the DSA – but also to the European Commission[100] – could ultimately meet the requirements of EU constitutional law and namely Article 52 (1) CFR. However, on closer inspection, there are significant differences between the DSM directive, to which the above judgment referred, and the DSA. This is particularly so with regard to the protection of the freedom of expression. In the DSM case, which ironically stems from an action for annulment brought by the Polish government,[101] the deliberately vague obligation to make "best efforts"[102] to prevent the upload of copyright-infringing content conflicted with the more concrete obligation that filtering should "not result"[103] in filtering out content that does not infringe copyright and related rights. Hence, an obligation of means (*obligation de moyens*) aiming to protect the rightsholders' right to intellectual property under Article 17 (2) CFR conflicted with an obligation of result (*obligation de résultat*) protecting the users' right to freedom of

[99] CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, para 75 (emphasis added), building on CJEU, Case C-314/12 *UPC Telekabel Wien* EU:C:2014:192, para 52.

[100] See, in particular, the best practices and guidelines to be established by the Commission under Article 35 DSA.

[101] Which is dismantling the rule of law and particularly judicial independence in a systematic way, see, in particular, CJEU, case C-619/18 *Commission v Poland [Independence of the Supreme Court]*, 2019; case C-192/18 *Commission v Poland [independence of ordinary courts]*, 2019; case C-585/18 *A.K. [Independence of the Disciplinary Chamber of the Polish Supreme Court]*, 2019; case C-824/18 *A.B. and Others [Appointment of Judges to the Polish Supreme Court]*, 2021; CJEU, case C-791/19 *Commission v Poland [Disciplinary Regime for Judges]*, 2021; CJEU, case C-487/19 *W.Ż. [Chamber of Extraordinary Control and Public Affairs]*, 2021; pending CJEU, case C-204/21 *Commission v Poland [independence and private life of judges]* and ECtHR, case No 4907/18 *Xero Flor*, 2021.

[102] Article 17 (4) Directive 2019/790, which in fact does not establish an obligation in the strict sense, but a condition for exempting service providers from liability. For reasons of technological neutrality, the EU legislator deliberately did not want to impose any precise requirements on the technical tools used by the providers.

[103] Article 17 (7) Directive 2019/790.

expression and information under Article 11 CFR.[104] Taking into consideration that a filter can hardly be qualified as "best effort" if it exposes the service provider to the real risk that lawful content is (systematically) blocked, the obligation of result tends to prevail over the obligation of means.[105] In simplified terms, the CJEU has construed the DSM Directive to be a framework within which the freedom of expression can be protected at a comparatively solid level because of the specific legislative design and the "appropriate safeguards by the EU legislature".[106]

By contrast, it is questionable whether the EU legislature has equally fulfilled this obligation of establishing appropriate safeguards when enacting the DSA. If one has struggled through the 156 recitals of the DSA, one might be ensouled by its noble fundamental rights ambitions. The legally binding part and particularly Article 14 (4) DSA, however, do not seem to do sufficient justice to these ambitions. Admittedly, there are procedural safeguards such as detailed obligations to give reasons[107] and to provide for transparency[108] as well as the provisions on notices,[109] complaint-handling[110] and dispute settlement.[111] In this respect, the DSA is certainly a step in the right direction. However, there are practically no specifications as to how the conflicting fundamental rights should be protected *in substantive terms*. Instead, the EU legislature grants providers an astonishingly broad discretion.[112]

This is all the more remarkable considering that the DSA also covers measures that are taken by providers with regard to content that is neither illegal under EU nor national law, but merely contradicts the providers' terms and conditions. This is a sensitive and complex area in which freedom of expression is particularly at risk. There may be good reasons for platforms – ethical, moral, political, economic – to delete or reduce the visibility of content that is, for instance, "lawful but awful". However, such measures should be guided by a clear legal framework that, while allowing the necessary latitude for balancing in individual cases, defines the legal conditions and limits for such

[104] Aptly Raue, 'Die Zähmung der Uploadfilter – Konsequenzen aus dem EuGH-Urteil Polen/Parlament und Rat zu Art. 17 DSM-RL', Zeitschrift für Urheber- und Medienrecht, 2022, 624, 628-629.

[105] Id., 629.

[106] CJEU, Case C-401/19 *Poland v Parliament and Council [Upload Filters]*, ECLI:EU:C:2022:297, para 98.

[107] Article 17 DSA.

[108] Articles 15, 24, 27, 39, 42 DSA.

[109] Articles 16, 18 DSA.

[110] Article 20 DSA.

[111] Article 21 DSA.

[112] Eifert et al., 'Taming the Giants: the DMA/DSA Package', 58 Common Market Law Review 2021, 987, 1013: 'unfettered discretion'.

measures. A legal act such as the DSA should give concrete expression to EU fundamental rights in the sense that it specifies both latitude *and limits* of the providers to shape their terms of content under Article 16 CFR. This holds true all the more since the DSA apparently aims at fully harmonising this aspect, with the result of precluding a further specification by means of national legislation.[113]Arguably, the EU legislature has failed in sufficiently giving concrete expression to EU fundamental rights in the above-mentioned sense, with perhaps the notable exception of the provisions on the protection of minors.[114] All in all, the DSA effectively amounts to an escape from legislative responsibility as far as the protection of EU fundamental rights is concerned. Assuming that the CJEU will not invalidate the DSA, effective judicial review becomes all the more important. This already leads to the aspect of enforcement.

4. Enforcement: The Crucial Role of National and European Courts

The DSA provides for different modes of enforcement. First of all, it establishes a special feature that only applies to the so-called "very large online platforms and very large online search engines".[115] These companies are obliged to carry out a risk assessment at the macro level in order to be able to prevent or to react to systemic dangers, stemming from the design, the functioning, or the use of the respective services.[116] The discretion given to the European Commission in this area with regard to setting best practices and guidelines is remarkable.[117]

Second, the DSA also provides for mechanisms of executive enforcement by the Member States. As we have already seen, the DSA largely leaves it to the Member States to determine the normative framework of executive enforcement. In doing so, the Member States can be bound both by EU and national fundamental rights.

Third, the DSA also establishes quasi-judicial mechanisms of enforcement at the individual or micro level. Article 20 DSA requires providers to ensure access to an effective "internal complaint-handling system". However, the DSA does not sufficiently specify if platforms in Article 20 DSA appeals processes have to apply fundamental rights or purely terms of services. Another forum envisaged by the DSA is "out-of-court" (OOC) dispute settlement, the purpose of which is to resolve disputes related to decisions of the internal complaint handling system (Article 21 DSA). Also in this regard

[113] Article 52 (1) CFR has also a federal dimension, as a national act can also fulfil the condition of being 'provided for by law', see CJEU, Case C-336/19 *Centraal Israëlitisch Consistorie van België and Others* ECLI:EU:C:2020:1031, para 60.

[114] Article 28 DSA.

[115] Cf. Article 33 DSA.

[116] Article 34 DSA.

[117] Cf. Article 35 DSA.

the DSA does not sufficiently clarify whether OOC bodies have to apply fundamental rights or purely terms of services. What is clear, however, is that OOC bodies are certified at the national level and do not have the power to impose a binding settlement of the dispute on the parties.[118] Otherwise, the decisions would have had to be mutually recognised as legally binding, which both the EU legislature and the service providers apparently wanted to avoid, not least in view of the rule of law crisis and the potential influence of autocratic governments on the certified dispute resolution bodies in several Member States. After all, who would like to be in the position of having to recognise a decision on freedom of expression issued by a dispute settlement body certified by Victor Orbán?

Interestingly, the prominent example of Facebook's (already existing) Oversight Board (OSB), is intended as an *independent* review mechanism[119] which complements the internal complaint-handling mechanisms. For instance, in a much-reported decision in 2021, the OSB upheld a decision by Facebook to restrict Donald Trump's access to posting content on his Facebook and Instagram accounts.[120] The "case law" of the OSB demonstrates very well that providers are already today exercising a quasi-judicial role which requires legal expertise.[121] But it also exemplifies that even (semi-)independent complaint-handling bodies do operate in a rather specific normative environment. The OSB does not only rely on human rights standards, which it draws particularly from the International Covenant on Civil and Political Rights (ICCPR), but also applies normative corporate standards, namely "Facebook's content policies" and "Facebook's values". This is precisely why it would have been desirable, if not mandatory under EU constitutional law, for the European legislature to have specified the requirements for providers arising from European fundamental rights and to have struck a balance between the conflicting fundamental rights already in the DSA.

The above considerations show that internal complaint management systems and out-of-court dispute resolution can at best complement judicial review, but by no means replace it. Ideally, they can be a quick and cost-effective means of already resolving some of the disputes that arise in connection with providers' decisions. However, it will be up to national courts, in dialogue with the CJEU, to ensure that *legal* standards and, above all, EU and national fundamental rights are fully respected. This is all the more important as the DSA does not spell out the substantive requirements to the extent that would have been necessary. In the absence of sufficient concretisation of the standards in the DSA, it is imperative that effective judicial review sets limits to the discretion of providers – limits that should have already been outlined in the DSA. Article

[118] Article 21 (2) DSA.

[119] However, it is funded through a trust set-up by Meta.

[120] The decision is available here: https://oversightboard.com/decision/FB-691QAMHJ.

[121] However, the OSB is composed in an interdisciplinary way.

21 (1) DSA makes it clear that judicial review must in no way be ruled out by the DSA. Every user retains the right to initiate proceedings, at any time, before a court under the applicable national or European law. This is in line with the case law of the CJEU, which has ruled that it must be ensured that the balancing of conflicting fundamental rights can be subject to effective judicial review.[122] And as we have also seen, the fact that the courts must ultimately strike a balance between conflicting fundamental rights in no way precludes the possibility for individuals to (directly) invoke these fundamental rights in horizontal disputes.[123]

IV. Conclusion

The DSA will be at the centre of intriguing and major legal cases in the years to come. The protection of fundamental rights in multipolar relationships and in the European multi-level system will undoubtedly play a major role in this regard. The paper showed that the DSA does not effect full harmonisation in all its parts, with the result of national fundamental rights still being of relevance in some areas, such as executive enforcement at the national level. Furthermore, the paper demonstrated that the DSA cannot, as a legislative act, establish a horizontal direct effect of EU fundamental rights, but that such effect is by no means excluded, taking into consideration the very nature and *telos* of several of the fundamental rights in question, in particular the freedom of expression and information and the right to privacy.

From a constitutional point of view, the decisive question appears to be less whether such horizontal effects are to be qualified as direct or indirect, but rather to what extent the EU legislature must predetermine the balancing of conflicting fundamental rights and thus limit the leeway of the providers and the reviewing courts in advance. A major weakness of the DSA lies in the fact that the EU legislature missed out on specifying the requirements for providers arising from European fundamental rights. May the European legislature in future no longer run from its legislative responsibility for fundamental rights, but take it up for the benefit of us all.

[122] CJEU, Case C-414/16 *Egenberger* ECLI:EU:C:2018:257, para 53.

[123] Ibid, para 80.

The Digital Services Act: A General Assessment*

Florence G'sell

I. Introduction

In late January 2023, about 38 technology companies, including major social media platforms such as YouTube, Facebook, Twitter, and TikTok, submitted a report to the EU Commission detailing their implementation of the EU's 2022 Code of Practice on Disinformation[1]. This code, agreed upon in June 2022, requires platforms to provide information on their approaches to combating online disinformation and foreign manipulation.[2] Notably, Elon Musk's Twitter was the only company to provide an incomplete report, "short of data, with no information on commitments to empower the fact-checking community".[3] Although the Disinformation Code is not legally binding, the Commission will link adherence to commitments made under it with compliance with the new Digital Services Act (DSA).[4] The submission of these reports is, therefore, an important initial evaluation of how tech companies will implement the DSA, which is expected to be one of the major challenges of the decade for major technology platforms.[5]

The Regulation (EU) 2022/2065, also known as the Digital Services Act (DSA), will soon be enforced. It was published in the Official Journal of the European Union on October 27, 2022, and came into effect on November 16, 2022. Online platforms were required to inform the Commission of the number of active end-users on their websites by February 17, 2023. Based on these user numbers, the Commission will assess

*Many thanks to Rachel Griffin, Ph.D. candidate at Sciences Po Law School, for her insightful feedback on a previous version of this piece.

[1] https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

[2] These reports are available on the Disinfocode.eu website, https://disinfocode.eu/.

[3] Goujard, 'Elon Musk's Twitter fails first EU disinformation test', Politico, 9. February 2023, https://www.politico.eu/article/elon-musk-twitter-fails-eu-first-disinformation-test-digital-services-act/.

[4] Lomas, 'Musk's Twitter gets "yellow card" for missing data in EU disinformation report', Tech Crunch, 9. February 2023, https://techcrunch.com/2023/02/09/elon-musk-twitter-eu-disinformation-code-report/.

[5] G'sell, 'A French Perspective on Elon Musk's Twitter', Lawfare, 3. January 2023, https://www.lawfareblog.com/french-perspective-elon-musks-twitter.

whether a platform should be classified as a very large online platform or search engine with more than 45 million users. Once designated as such, these platforms will have four months to comply with their obligations under the DSA. This includes providing the Commission with its first annual risk assessment report. Apart from very large providers, all regulated entities must comply with the DSA fifteen months after it enters into force starting from February 17, 2024. In France, platforms are already subject to binding obligations included in the DSA, since the French Parliament passed Law No. 2021-1109[6] on August 24, 2021, which created a large number of provisions designed to transpose the DSA in advance. These provisions will sunset on December 31, 2023.

The DSA introduces a new regulatory framework for online platforms. Its goal is to encourage them to fight objectionable content while respecting users' fundamental rights. To that end, the DSA updates and complements the provisions of Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the internal market (e-commerce Directive), since this Directive no longer appears adequate for governing today's platforms which operate globally, are predominantly managed by algorithmic systems, and host content that may be harmful.[7] The adoption of the DSA is a significant achievement resulting from a long-term effort by European authorities to promote responsible moderation practices among major social media platforms which increasingly rely on automated tools to curate and moderate content.[8] Over the past two decades, as social networks have expanded, the trend towards the privatization and automation of online speech control has raised concerns.[9] Flawed moderation practices have prompted European authorities to take action and encourage platforms to implement effective policies against online hate and disinformation. In May 2016, Facebook, Microsoft, Twitter, and YouTube entered into an agreement with the EU Commission known as the "Code of Conduct on countering illegal hate speech online"[10] to prevent and combat the spread of hate speech on their respective platforms. Over time, other tech companies have also joined this code of conduct. In 2018, the EU Commission also introduced the aforementioned "Code

[6] Loi n°2021-1109 du 24 août 2021 confortant le respect des principes de la République, https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000042635616/.

[7] G'sell, 'Les réseaux sociaux, entre encadrement et auto-régulation', Sciences Po, Chaire Digital, Gouvernance et Souveraineté, november 2021, https://www.sciencespo.fr/public/chaire-numerique/2021/11/09/research-paper-les-reseaux-sociaux-entre-encadrement-et-auto-regulation/.

[8] Narayanan, 'Understanding Social Media Recommendation Algorithms', Knight First Amendment Institute at Columbia University, 2023.

[9] Balkin, 'Free Speech is a Triangle', Columbia Law Review, vol. 118: 2011, 2018; Klonick, The New Governors: the People, Rules, and Processes Governing Online Speech, Harvard Law Review, vol.131: 1598, 2018.

[10] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

of Practice on Disinformation"[11], which around 38 tech companies have committed to following. This code was amended and strengthened in 2022 and includes a series of commitments and specific measures designed to address concerns related to disinformation. Finally, on March 1, 2018, the EU Commission published its "Recommendation C/2018/1177 on Measures to Effectively Tackle Illegal Content Online"[12], which encouraged tech companies to enhance their notice and action procedures, among other things, to more effectively address illegal content on their platforms.

However, the DSA is not the only recent legislative instrument affecting platforms' liability and content moderation policies. Two legislations, in particular, have consistently strengthened the liability of online platforms and increased their obligations. The Directive 2019/790 on copyright and related rights in the Digital Single Market[13] establishes that providers of online content-sharing services are directly responsible when users illegally upload protected content. Providers may be exempt from liability if they have made best efforts to obtain authorization from the rightsholder and to block unauthorized content if they acted expeditiously to remove content following a notification from a rightsholder and if they proactively prevented future upload of that content. Regulation 2021/784 of 29 April 2021[14] on combating terrorist content online requires hosting service providers to take measures to prevent its dissemination, including removing terrorist content within one hour after receiving a notice from law enforcement. In this context, the adoption of the DSA is another evolution in the EU's ongoing efforts to regulate online platforms and fight against illegal activities on the internet.

The purpose of this article is to provide a brief overview of the Digital Services Act and offer commentary on its main features. The article is structured as follows. The first part presents the key features of the DSA, which can be summarized in five points: the asymmetric character of the DSA, which adapts the rules and obligations to the size and activities of the regulated entities, the preservation of the liability exemption established by the E-Commerce Directive, the creation of new due diligence obligations regarding content moderation, the implementation of specific obligations designed to protect users and consumers and a crisis mechanism, and the implementation of the DSA. The second part of the article aims to identify potential implementation difficulties, including those associated with enforcing the DSA, managing systemic risk, and applying the DSA to emerging technologies.

[11] https://disinfocode.eu/.

[12] https://digital-strategy.ec.europa.eu/en/library/commission-recommendation-measures-effectively-tackle-illegal-content-online.

[13] https://eur-lex.europa.eu/eli/dir/2019/790/oj.

[14] https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv%3AOJ.L_.2021.172.01.0079.01.FRA.

II. Key Features of the DSA

The Digital Services Act has a vast and comprehensive scope that aims to regulate the activities of "intermediary services" offering digital services to legal entities in the European Union, as stated in Article 2. It is immaterial whether the provider is based in the EU or not under the DSA. If the service provider does not have an establishment in the EU, the Regulation's applicability is subject to the "substantial connection" condition with the EU, as outlined by Article 3(d). This means that the provider must have a significant number of EU-based users or target its activities towards a specific EU member state, such as by using a relevant top-level domain name (Article 3(e) DSA). All providers of intermediary services must appoint a single point of contact allowing for direct communication with the competent supervisory authorities and users, as provided by Articles 11 and 13.

The key features of the Digital Services Act can be summarized in five points. First, it is an asymmetrical regulation. Second, it upholds the principle of exemption from liability while introducing the Good Samaritan principle. Third, it introduces new obligations for content moderation to combat undesirable content effectively and better protect users' rights. Fourth, it includes specific provisions aimed at enhancing user and consumer protection. Finally, it contains very specific implementation and enforcement procedures.

1. Asymmetric Regulation

While the scope of the Digital Services Act is broad, it solely governs "intermediary services" rather than the "information society services" regulated by the E-Commerce Directive. "Intermediary services" refer to those that transmit and store user-generated content, as per Article (3)(g) DSA. The DSA further classifies different types of "intermediary services" that are available to users in the European Union. The first three categories were already present in the E-commerce Directive and include "mere conduit"

services,[15] "caching" services,[16] and "hosting" services.[17] The DSA adds two new categories: "online platforms", which are a category of hosting services that disseminate information to the public,[18] and "online search engines".[19]

The distinctions among different service types are significant because the DSA is not meant to be uniformly applied to all regulated service providers. Instead, the DSA takes the form of a "layer cake", designed to be applied asymmetrically with rules that vary depending on provider characteristics. In other words, the DSA's obligations are structured as a pyramid,[20] with layered requirements from the bottom to the top. At the base of the pyramid, the first layer encompasses all intermediary services that have very basic obligations, followed by hosting services, and then online platforms. Moving up the pyramid reveals increasingly stringent obligations that apply to fewer and fewer categories of providers. At the top of the pyramid, the most extensive and restrictive obligations are imposed on very large online platforms (VLOPs) or search engines (VLOSEs) that have at least 45 million average monthly active users in the EU. These additional obligations are justified by the systemic risks they pose due to their size.

While large companies face heavier obligations, micro and small companies are exempt from certain obligations. For instance, transparency obligations (Article 15), provisions applicable to online platforms (Section 3), and provisions applicable to platforms allowing consumers to conclude distance contracts with traders (Section 4) are not applicable to micro or small enterprises. These small enterprises are defined as companies with fewer than 250 employees and an annual turnover under €50 million or an annual balance sheet total under €43 million, as per Recommendation 2003/361/EC.[21] Despite this exemption, it could be argued that the threshold is too low and that the DSA's stringent obligations may negatively impact the financial stability and growth of small and medium-sized enterprises since the companies that are just

[15] That consist of the transmission of information on a communication network, or the provision of access to a communication network, as Internet service providers (article 3(g)(i) DSA).

[16] That consist of the transmission of information involving the automatic, intermediate, and temporary storage of that information, as web browsers (article 3(g)(ii) DSA).

[17] That consist of the storage of information provided by users (article 3(g)(iii) DSA).

[18] Which are hosting services that store and disseminate information to the public, as social networks or marketplaces (article 3(i) DSA).

[19] That allow users to input queries in order to perform searches on the basis of a query on any subject (article 3(j) DSA).

[20] Wilman, 'The DSA, an overview', Nederlands tijdschrift voor Europees recht, No. 9/10, 2022, p. 220, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4304586; Husovec and Roche Laguna, 'Principles of the Digital Services Act', Oxford University Press, 2023; Wilman, 'The responsibility of online intermediaries for illegal user content in the EU and the US', Cheltenham: Edward Elgar Publishing 2020; Wilman, 'The EU's system of knowledge-based liability', 12, 2021, JIPITEC 317.

[21] Recommendation of 6. May 2003, Article 2, Annex 1.

above these thresholds would be penalized.[22] As a result, the DSA would benefit larger platforms and search engines that have the resources to comply with its provisions. In any case, an assessment of the DSA's impact on micro and small companies will be conducted by the Commission after five years, as provided by Article 91(2)(d).

2. Liability Exemption

The DSA preserves the exemption from liability established by the E-Commerce Directive in 2000, with additional clarifications. One significant addition is the inclusion of the Good Samaritan clause, which draws inspiration from section 230 of the US Communications Act of 1934.

a) Preservation of the Liability Exemption

Before the implementation of EU Directive 2000/31/EC, it was not clear to what extent service providers were liable for the content posted on their platforms. The E-Commerce Directive then introduced a new category of service providers called "hosting service providers" and provided, in its Article 14, that these hosting providers are exempt from liability for content stored at a user's request if they have no actual knowledge of its illegality. Providers of mere conduit and caching services are similarly not held liable for the information they transmit or store for their users.

The DSA upholds the liability exemption provided by the E-Commerce Directive for over 20 years. Article 6 of the DSA grants hosting providers immunity from being held liable for any illicit content that may be present on their platforms. However, this protection only applies if they act "expeditiously" to remove access to the content once they become aware of its illegality. Similarly, Article 4 specifies that mere conduit service providers are not liable for the information transmitted or accessed if they do not initiate the transmission, select the receiver of the transmission, or modify the information contained in the transmission. Under Article 5, caching service providers are not liable if they do not modify the information and act expeditiously to remove or disable access to the information upon obtaining actual knowledge of the fact that the information has been removed from its initial source or when required by law or a court order. They must also not interfere with the lawful use of technology. Unlike the provisions of the E-Commerce Directive, which had to be transposed into national law, this liability exemption now applies directly and uniformly across all EU countries.

[22] See Keller, 'The DSA's Industrial Model for Content Moderation', Verfassungsblog, 24. February 2022, https://verfassungsblog.de/dsa-industrial-model/; Keller, 'The EU's new Digital Services Act and the Rest of the World', Verfassungsblog, 7. November 2022, https://verfassungsblog.de/dsa-rest-of-world.

In addition to upholding the liability exemption, the DSA provides clarifications that were not included in the E-Commerce Directive, often based on existing case law. For example, Article 16 (3) presumes actual knowledge if the content has been reported and its illegality is apparent without detailed legal examination. This was already the case in France where the Constitutional Council ruled that hosting providers can only be held liable if they remain inactive while aware of "manifestly illicit" content.[23] Furthermore, the exemption does not apply if the intermediary service provider, "instead of confining itself to providing the services neutrally by a merely technical and automatic processing of the information" plays an active role that gives them knowledge or control over the content, as stated by Recital 18, codifying the CJEU caselaw.[24] However, as algorithmic recommendation systems developed by platforms increasingly determine the organization of content presented to users, some may question whether the traditional differentiation between active and passive roles remains relevant. Despite this objection, the DSA maintains the distinction and provides, in Recital 22, that actual knowledge doesn't result from the fact that the platform automatically indexes information, offers a search function, or recommends information on the basis of the profiles or preferences of users.

b) Introduction of the Good Samaritan Principle

Article 7 of the DSA allows providers to carry out voluntary investigations or take other measures to detect, identify, and remove illegal content. However, some may worry that this could lead to platforms being considered to play an active role that gives them knowledge or control over the content and losing their exemption from liability as a result. In reality, engaging in these investigations does not automatically make platforms responsible for the content. The Good Samaritan clause was added to the DSA in response to requests from online platforms for greater clarity and reassurance that they could take voluntary steps to remove illegal content without losing their liability exemption.

The newly introduced Good Samaritan clause draws inspiration from Section 230 of the US Communications Act of 1934,[25] specifically 230(c)(2)(A), which offers Good Samaritan immunity to platforms. This immunity allows platforms to intervene in good faith on content without incurring any liability. In the physical space, the Good Samaritan immunity protects those who try, in good faith and without legal obligation,

[23] Decision n° 2004-496 DC of 10. June 2004, https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2004-496-dc-du-10-juin-2004-communique-de-presse.

[24] Case C-324/09, *L'Oréal v eBay* CJEU 12 July 2011; joined cases C-682/18 (*Cyando*) and C-683/18, (*YouTube*) CJEU 22. June 2021.

[25] 47 U.S. Code § 230.

to assist individuals in distress. In the virtual space, the Good Samaritan immunity guarantees that providers and users of online services will not be held liable for any action taken in good faith to remove or restrict access to content that the provider or user considers objectionable. Under the DSA, the Good Samaritan clause ensures that providers can detect and remove illegal content without losing their liability exemption.

Article 7 of the DSA, following the example of Section 230, requires providers to act in good faith and with due diligence in their effort to remove illegal content, thereby limiting the scope of the Good Samaritan immunity. It is unclear, however, whether these requirements are adequate in preventing the risk of encouraging excessive monitoring or moderation of content by platforms. To mitigate these risks, Article 7 and Recital 26 of the DSA provide important guidance on the measures that can be taken to detect illegal content. Providers must not only act in good faith and diligently but also take objective, non-discriminatory, and proportionate measures. In addition, they must provide safeguards against the unjustified removal of legal content. Recital 26 specifies that providers using automated tools should "take reasonable measures" to ensure that "the technology is sufficiently reliable to limit to the maximum extent possible the rate of error". It is a positive development that the DSA introduces the consideration of the error rate of algorithmic models, but its implementation should certainly go beyond this rather vague recommendation.

3. New Due Diligence Obligations for Content Moderation

The DSA introduces additional obligations for intermediary service providers beyond the relative exemption from liability that has been in place for over two decades. The new obligations provided by the DSA stem from concerns about the effectiveness of the existing knowledge-based liability principle in compelling platforms to address illegal content. Ironically, these new obligations also result from the concern that the current principle may lead platforms to over-censor and remove excessive amounts of content. Therefore, these obligations are aimed at ensuring hosting providers combat undesirable content effectively, while safeguarding users' fundamental rights, particularly freedom of expression. Non-compliance with these obligations not only exposes providers to sanctions by regulatory authorities but also entitles affected users to seek compensation for any damages suffered, as provided by Article 54.

The DSA's new obligations are primarily centered around the content moderation activities carried out by the platforms. This is why the DSA introduces, in Article 3(t), a broad definition of "content moderation" as "the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at detecting, identifying, and addressing illegal content or information incompatible" with the providers' terms and conditions, including "measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such

as demotion, demonetisation, disabling of access to, or removal", or that affect the ability of users to publish or transmit information, such as the termination or suspension of a user's account.[26] With this definition, the DSA recognizes the crucial role played by platforms in moderating content, frequently through the use of automated tools. It also acknowledges that this moderation activity is not only based on the applicable laws but is also governed by the platforms' terms and conditions, which is reflected by the fact that most major platforms now publish increasingly detailed content policies.

Consequently, the DSA's new obligations related to content moderation can be grouped into four categories: combating illegal content, upholding procedural safeguards in moderation, ensuring transparency, and managing systemic risks.

a) Combating Illegal Content

The DSA not only maintains the liability exemption, but it also imposes new obligations on online platforms to effectively combat illegal content. This fight against illegal content involves the users and relies mainly on user notifications. The DSA provides comprehensive guidelines on how to handle user notifications, which have not been as detailed in the past.

Article 16 requires providers of hosting services to establish accessible "notice and action" mechanisms that allow anyone to notify them of the presence of illegal content. The main goal is to make sure that these notice and action procedures are effective in combating illegal content while also safeguarding the rights of users, including protection against unjustified removal. Therefore, hosting platforms must implement efficient reporting mechanisms and have clear and user-friendly reporting systems in place to enable users to report illegal content. To that end, providers are required by Article 12 to designate a single point of contact to facilitate direct and rapid communication through electronic means. In addition, Article 22 requires providers to prioritize reports from trusted flaggers, which are entities designated by competent national authorities that have demonstrated expertise and competence in identifying illegal content. Hosting service providers must process received notifications in a timely, diligent, non-arbitrary, and objective manner, as provided by Article 16(6). However, it must also be highlighted that providers are not obligated to take action on the reported content following the reception of a user's notice. They are expected to remove the content only when the notice is sufficiently clear and adequately substantiated, and when the illegality can be established without a detailed legal examination, as outlined by Article 16(3).

[26] Regarding the various content moderation actions that providers can take, see Goldman, 'Content Moderation Remedies', 28 Michigan Technology Law Review 1, 2021, https://repository.law.umich.edu/mtlr/vol28/iss1/2.

The fight against illegal content also depends on effective collaboration with national authorities. Article 18 imposes an obligation on online platforms to cooperate with national law enforcement or judicial authorities and promptly notify them of any content that may give rise to suspicions of criminal offenses posing a threat to the life or security of individuals. The purpose of this obligation is to prevent or swiftly address serious crimes. However, this obligation is limited to criminal offenses that pose a threat to the life or safety of one or more persons, and it does not cover other criminal acts. Additionally, providers must comply with any instructions from authorities to act against illegal content and must justify the measures taken. Articles 9 and 10 provide that national judicial or administrative authorities may issue orders requiring providers to act against specific illegal content or provide information about certain users, but such orders must not constitute a general monitoring obligation.

Indeed, the DSA maintains, in Article 8, the prohibition of mandated general monitoring that already existed in the E-Commerce Directive. The prohibition concerns obligations "of a general nature" as opposed to obligations "in a specific case", as stated by Recital 30. An example of an obligation "of a general nature" is the obligation to introduce a system for filtering all electronic communications for an unlimited period and at the provider's expense in order to block unlawful use or transfer of copyrighted works.[27] However, an obligation for a service provider to identify and remove specific information deemed illegal by a court and equivalent information is not covered by the prohibition.[28] Despite existing precedents, it is not always easy to distinguish between general and specific obligations.[29]

b) Upholding Procedural Safeguard in Content Moderation

Online platform providers not only moderate content in accordance with the relevant laws and regulations but also set their terms and conditions, which allow them to determine their own content policies and decide what content to host or remove. These moderation standards serve as a private norm that governs online speech in practice. The DSA is actually one of the first pieces of legislation to recognize the crucial role played by terms and conditions in content moderation, while also trying to guarantee that the determination of these standards and their enforcement is in accordance with

[27] Case C-70/10, *Scarlet Extended v SABAM* (CJEU 24. November 2011). Article 17 of the Copyright Directive now obliges platforms to automatically filter copyright-infringing content, see C-401/19 *Poland v Parliament and Council*, (ECJ, 26. April 2022).

[28] Case C-18/18, *Glawischnig-Piesczek v Facebook Ireland Limited* (CJEU 3. October 2019).

[29] Kuczerawy, 'General Monitoring Obligations: A New Cornerstone of Internet Regulation in the EU?', in Rethinking IT and IP Law - Celebrating 30 years CiTiP, Intersentia, 2019, https://ssrn.com/abstract=3449170.

the fundamental rights of users. To that end, the DSA adopts a procedural perspective and imposes due process obligations intended to serve as safeguards against possible arbitrariness by platforms. Furthermore, the DSA does not interfere with the freedom of hosting providers to establish their own content policies.

Some authors argue that prioritizing "procedure over substance"[30] is insufficient in effectively safeguarding users' rights, as these procedures may only be understood and utilized by the most informed or privileged individuals.[31] However, procedural fairness has been demonstrated to be an effective method of defending fundamental rights, and it could be argued that it is the most pragmatic approach to limiting the overwhelming power and potential arbitrariness of platforms by providing quasi-constitutional guarantees.[32] The approach favored by the DSA could actually be compared to the judicial process since it requires that each moderation decision be based on a known pre-existing rule, motivated by arguments, and open to challenge by the affected user.[33]

The DSA stipulates in Article 14 that providers are required to clarify "any restrictions that they impose in relation to the use of their service" in their terms and conditions. The information provided must include "any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making, and human review as well as rules of procedure of their internal complaint handling system". This requirement could be analyzed as a form of "codification" of moderation rules,[34] enabling the enforcement of "the rule of law's principles of legality, predictability, and accessibility for the imposition of sanctions".[35] Furthermore, Article 14(4) mandates that service providers must act diligently, objectively, and proportionately while applying and enforcing their terms and conditions, taking into account the rights and legitimate interests of all parties involved, including users' fundamental

[30] Ortolani, 'If You Build It, They Will Come', Verfassungsblog, 7. November 2022, https://verfassungsblog.de/dsa-build-it/.

[31] Griffin, 'Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality', European Law Open, 2023, 2 (1), forthcoming.

[32] Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms', 2018, 4 (3) Social Media + Society 4, https://doi.org/10.1177/2056305118787812; De Gregorio, 'Democratising online content moderation: A constitutional framework', 2019, 36, Computer Law & Security Review 105374, https://doi.org/10.1016/j.clsr.2019.105374.

[33] Douek, 'Content Moderation as Systems Thinking', 2022, 136, Harvard Law Review 526, https://harvardlawreview.org/2022/12/content-moderation-as-systems-thinking/.

[34] Leerssen, 'An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation', Computer Law and Security Review, 48, 2023, 105790, https://www.sciencedirect.com/science/article/pii/S0267364923000018.

[35] Ibid.

rights.[36] It is a plausible assumption that users could rely on this provision to file a liability claim in situations where they feel their rights have been violated, even if they have not been affected by any moderation decision. This could occur when a platform's decision deprives them of access to information they are interested in, for instance when a politician's or journalist's account is suspended,[37] or when a social network abruptly decides to ban parody accounts,[38] or when it becomes evident that a platform systematically discriminates against content from a particular group of people.[39]

In addition to clarifying their terms of use, platforms are obliged to provide a clear and specific explanation of the reasons for the decisions they make about any content provided by users constituting either illegal content or being incompatible with their terms and conditions. Users will this be able to challenge moderation decisions made about them, thanks to the "statement of reasons" they receive. The scope of the obligation to provide an explanation is particularly broad in this regard. Indeed, as illustrated by the wording of the abovementioned Article 3(t), the DSA's definition of moderation actions is comprehensive and all-encompassing. Article 17 stipulates an explanation must be provided in the case of "any restrictions of the visibility" of content, including "removal of content, disabling access to content, or demoting content", and in case of suspension, termination, or other restriction of monetary payments, of the service, or of the service's account. In other words, a statement of reasons is required in cases where content is removed, demonetized, or demoted, including instances of "shadow banning," where a user's content is concealed from others without their knowledge.[40]

According to Article 20, providers must establish an efficient internal complaint-handling system that allows users to dispute moderation decisions they believe to be unjust or incorrect. This system should enable users to contest decisions related to the removal of allegedly illegal content or the suspension of their account or service provision, as well as decisions not to act on a notice of illegal content. The DSA provides that such complaints must be reviewed "in a timely, non-discriminatory, diligent, and non-

[36] Appelman, Quintais, Fahy, 'Using terms and conditions to apply fundamental rights to content moderation: is article 12 DSA a paper tiger?', Verfassungsblog, 10. September 2021, https://verfassungsblog.de/power-dsa-dma-06/.

[37] Which happened when Donald Trump's account was suspended or when Elon Musk abruptly decided to suspend the accounts of journalists he accused of following him and disclosing his movements: Grynbaum, 'Elon Musk Flexes His Media Muscle by Suspending Reporters on Twitter', The New York Times, 16. December 2022, https://www.nytimes.com/2022/12/16/business/media/elon-musk-twitter-journalist-suspension.html.

[38] 'Elon Musk's first Twitter moderation change calls for permanent bans on impersonators', The Verge, 7. November 2022, https://www.theverge.com/2022/11/6/23443871/elon-musk-twitter-perma-ban-impersonation-parody.

[39] For instance, conservatives feel that they are subject to unjustified censorship.

[40] See Leerssen, 'An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation', https://www.sciencedirect.com/science/article/pii/S0267364923000018.

arbitrary manner", as stated by Article 20(4). Moderation decisions must be reversed when the complaint contains sufficient grounds. It is important to emphasize that automation does not appear to be entirely excluded in the decision-making process for ruling on the complaint. According to Article 20(6), the appeal decisions must be "taken under the supervision of appropriately qualified staff, and not solely on the basis of automated means". This provision lacks clarity and one might be concerned about its practical implications. This seems to suggest that platforms, which may have limited resources, could rely on algorithmic methods to process appeals. Appeal decisions would then be quickly reviewed and validated by a human. In this scenario, the likelihood of successful appeals overturning moderation decisions appears to be relatively low.

Users also have the option to submit disputes to an out-of-court dispute settlement body certified in one of the Member States based on their independence and expertise, as per Article 21. However, this provision may not be effective since Article 21(3), subparagraph 3 provides that "the certified out-of-court dispute settlement body shall not have the power to impose a binding settlement of the dispute on the parties". This means that in case of persistent disagreement on moderation decisions, users will have no other option than to go to court at their own expense, as provided by Article 21(1), subparagraph 3. Indeed, Article 54 provides that users can always request compensation for any damages or losses incurred due to a breach of the DSA. They can even bring representative actions for the protection of collective consumer interests, as per Article 90.

The insistence on safeguarding users' rights does not exclude that users can be sanctioned when they repeatedly adopt behaviors contrary to the law or the platform's terms and conditions. The DSA includes provisions to sanction users who repeatedly publish manifestly illegal content or submit manifestly unfounded notices or complaints through notice and action mechanisms. Article 23 provides that online platform providers are required to suspend services to such users, subject to several safeguards. The user must have been previously warned and the suspension must also be limited to a reasonable period of time. Providers must assess each case individually, in a timely, diligent, and objective manner, and clarify their policies in advance in the terms and conditions.

c) Ensuring Transparency

The DSA not only implements procedural due process but also seeks to incentivize platforms to act in a responsible manner by imposing precise transparency obligations on them, especially regarding their moderation practices. The objective is to provide information to the general public, media, experts, and regulators on information that is currently challenging to obtain like, for instance, information regarding the resources

allocated to moderation, which Twitter was ordered to disclose in France.[41] Additionally, there is a vast amount of information related to the actual functioning of moderation, such as the quantity of removed content or the error rate of algorithmic models utilized in moderation, that remains inaccessible.

According to Article 15 of the DSA, all intermediaries except micro and small enterprises must report annually on their content moderation. The report must provide information about content moderation at the providers' own initiative, including the use of automated tools. It must include, among other things, the number and type of measures taken that affect information availability, visibility, and accessibility and the number of orders received from Member States' authorities categorized by the type of illegal content concerned. Hosting providers must also disclose "the number of notices submitted categorized by the type of alleged illegal content concerned, the number of notices submitted by trusted flaggers, any action taken pursuant to the notices by differentiating whether the action was taken on the basis of the law or the terms and conditions of the provider, the number of notices processed by using automated means and the median time needed for taking the action" (Article 15(1)(b)). Online platforms must add information about the basis for the complaints received, the decisions taken following those complaints, the median time needed for making those decisions, and the number of instances where those decisions were reversed. They must also provide information about the number of disputes submitted to out-of-court settlement bodies and the number of service suspensions following the publication of manifestly illegal content or manifestly unfounded notices or complaints, as provided by Article 24. The information reported must be categorized by the type of illegal content or violation of the terms and conditions of the service provider, by the detection method, and by the type of restriction applied.

The DSA imposes increased transparency and accountability measures for very large online platforms (VLOPs) and search engines (VLOSEs). These operators must publish transparency reports twice a year that include information about their content moderation resources and the qualifications of their moderators, as provided by Article

[41] In France, Twitter was sued for its failure to cooperate with judicial authorities in moderating hateful speech on its platform: Laurent et Leloup, "Twitter poursuivi en France pour son manque de cooperation dans des affaires de haine en ligne", Le Monde, 3. February 2021, https://www.lemonde.fr/societe/article/2021/02/03/haine-en-ligne-twitter-poursuivi-en-france-pour-son-manque-de-cooperation-avec-les-services-de-police_6068567_3224.html; On 6. July 2021, the Paris Court (TJ Paris, 6 juillet 2021, n°°20/35181) ordered Twitter to disclose details about the resources it employed to combat hate speech, including both human and material resources, https://cdnx.nextinpact.com/data-next/file-uploads/minute__1_.pdf; This ruling was subsequently upheld by the Paris Court of Appeal (CA Paris, 20 janvier 2022, n°21/14325) and the Cour de cassation (Cass. Première présidence, ordonnance, 23 mars 2023, n°22-13.600).

42. In addition, VLOPs and VLOSEs must provide regulators with access to the data needed to verify that they are in compliance with the DSA, as stated by Article 40(1)). In this respect, they must be able to explain to regulators the design, logic, operation, and testing of their algorithmic systems, including their recommendation systems.

While some authors have expressed disappointment that the transparency obligations do not require more detailed information to be disclosed,[42] the requirements of the DSA will still result in the disclosure of essential information that will enhance understanding of platform moderation practices. Of particular note is the fact that the DSA explicitly stipulates that the operation of automated tools must be specified since article 15(1)(e) requires "a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error" of the automated tools used.

d) Managing Systemic Risks

Article 34 of the DSA provides that very large online platforms (VLOPs) and search engines (VLOSEs) with at least 45 million users must evaluate and address "systemic risks" through appropriate policies. They must analyze the extent to which their moderation, recommendation, and advertising systems may affect those systemic risks. This should be done annually and also prior to the deployment of functionalities that are likely to have a critical impact on the risks identified.

Systemic risks pertain to issues such as illegal content, hate speech, privacy violations, election manipulation, and other similar problems. Moreover, content that generates adverse effects on fundamental rights, civic discourse, electoral processes, public security, gender-based violence, public health, minors, and personal well-being may also lead to systemic risks. Although Article 35(1) mentions "illegal hate speech or cyber violence", the definition of systemic risks encompasses content that is not necessarily illegal but may cause problems, such as misinformation on public health, climate change, or politics.

After assessing systemic risks, VLOPs and VLOSEs should implement "reasonable, proportionate, and effective mitigation measures" to counter such risks, as provided by Article 35. These measures may include adapting the design of the interfaces, adapting the terms and conditions, improving the notice and action mechanism, improving the algorithmic systems, increasing the visibility of reliable information sources, labelling suspicious content, or implementing codes of conduct, as provided by Article 35(1). The text, which refers to the possibility of "adapting any relevant decision-making processes and dedicated resources for content moderation" actually opens up many possibilities.

[42] Griffin, 'Rethinking Rights in Social Media Governance', fn. 32.

Moreover, the DSA introduces a unique monitoring system to enforce compliance with these obligations, which includes vetted researchers in addition to national and European regulatory bodies. First, VLOPs and VLOSEs are required to provide their assessments of systemic risks to the European Commission and relevant Digital Services Coordinators upon request, as per Article 35(2). The European Board of Digital Services will work with the Commission to publish annual reports on the identification and assessment of systemic risks, including best practices for mitigating these risks (Article 35(2)). Additionally, the Commission may issue guidelines and recommend actions in cooperation with Digital Services Coordinators (Article 35(3)). Within this framework, regulators play a role in determining and approving the strategies implemented to mitigate systemic risks. Second, regulators will also benefit from the expertise of researchers who will be provided with access to platform data to evaluate systemic risks. Indeed, Article 40 states that VLOPs and VLOSEs must provide internal data on request to researchers vetted by national regulators. This provision allows researchers to request whatever data they need to assess these risks and to go much further than the analyses provided in the reports prepared by the platforms themselves.

By providing for measures to address the "systemic risks" generated by the operation of large platforms and search engines, the DSA goes far beyond the simple fight against content deemed illegal by the national laws of the various Member States. This includes considering not only illegal content but also "lawful but awful" content that may be harmful.[43] In particular, while European national laws often criminalize hate speech, the same cannot be said of disinformation, which is often difficult to define or demonstrate and is rarely sanctioned as such.[44] In this context, the category of systemic risks can ideally serve as a basis for implementing effective policies to fight disinformation, in line with the Code of Practice on Disinformation[45] implemented at the European Union level. Furthermore, Article 35 enables regulators to issue guidelines and recommendations to mitigate systemic risks, which could include suggestions on the content of platforms' terms and conditions and content policies. Although the regulators' stance is yet to be determined, it is worth noting that such an opportunity is unprecedented.

[43] Keller, 'Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users', The University of Chicago Law Review Online, 28. June 2022, https://lawreviewblog.uchicago.edu/2022/06/28/keller-control-over-speech/.

[44] About French law, see G'sell, 'A French Perspective on Elon Musk's Twitter', https://www.lawfareblog.com/french-perspective-elon-musks-twitter, fn. 5.

[45] See fn 1.

4. Other Obligations

In addition to regulating moderation practices, the DSA contains provisions designed to protect users of online services more generally and in particular consumers who use the services of marketplaces and collaborative economy platforms.

a) Specific Protection of Online Platform Users

The DSA includes specific provisions for recommendation systems. According to Article 27, online platforms must provide precise and intelligible information in their terms and conditions about the main parameters used by their recommendation systems. This information must include the reasons for the relative weight of these parameters and the criterion with the most weight in selecting information presented to users. If multiple recommendation system settings are available, providers must allow users to select and change their preferred option at any time as per Article 27(3). This functionality must be directly and easily accessible. In addition, VLOPs and VLOSEs must offer users the option to choose a recommendation system that does not rely on profiling,[46] as provided in Article 38. This allows users to make informed decisions about their online experience and ensures their autonomy is respected.

Furthermore, the DSA strictly regulates online advertising. Advertising platforms must fully disclose their practices and targeting methods to advertisement recipients as per Article 26. Online platform providers must indicate on whose behalf and why the advertisement is being displayed to the user (Art. 26(1) DSA). VLOPs and VLOSEs must maintain a publicly accessible repository containing information about advertisements presented, including their content and the companies on whose behalf they were made, as provided by Article 39 DSA. Additionally, Article 26(3) prohibits displaying advertising based on profiling using sensitive data such as political opinions, religious beliefs, and sexual orientation. Targeted advertising of minors based on their personal data is prohibited, and specific protection measures must be put in place to ensure their safety online, as per Article 28.

Finally, the DSA strictly prohibits the use of "dark patterns" that manipulate internet users into performing a specific action, such as subscribing to a service, by subconsciously influencing them. According to Article 25(1), providers of online platforms cannot "design, organise or operate their online interfaces in a way that deceives or manipulates" users or in a way that "otherwise materially distorts or impairs the ability of" users "to make free and informed decisions". This prohibition may be difficult to implement in practice due to its vague wording. It does not apply when other regulations, such as GDPR, are in effect.

[46] As defined by Article 4 (4) of GDPR.

b) Specific Protection of Consumers on Marketplaces and Collaborative Economy Platforms

Online platforms that allow the conclusion of distance contracts between consumers and traders, such as marketplaces and collaborative economy platforms, have specific obligations that they must fulfill. These obligations include obtaining certain information about their professional users, such as their name, contact details, and identification and registration information, through "know your customer" protocols. Article 30 provides that online platforms must make best efforts to verify the accuracy and completeness of the information provided by professional users. In addition, these platforms are required by Article 31 to ensure that their interface is designed in a way that complies with consumer law regarding pre-contractual information obligations and product safety.

Additionally, these platforms are required to make "reasonable efforts to randomly check in any official, freely accessible and machine-readable online database or online interface whether the products or services offered have been identified as illegal", as per Article 31(3). If they become aware of any such illegality, they must inform the consumers who purchased the illegal product or service about the trader's identity and all relevant means of redress, as provided by Article 32 DSA.

Finally, Article 6(3) states that online platforms allowing consumers to enter into distance contracts with traders are not exempt from liability under consumer protection law in certain cases. The liability exemption does not apply when the platform presents information or enables a transaction in a way that leads an average consumer to believe that the information or product/service is provided by the platform itself or by a service recipient acting under its authority/control. In this respect, the DSA represents a significant evolution from the E-Commerce Directive. Previously, providers were only liable if it was proven that the offending professional user acted under their authority/control, as provided by Article 14(2) Directive 2000/31/EC. Now, liability is based on the consumer's perspective and what they may have reasonably thought. This makes sense since these provisions do not seem to intend to hold marketplaces liable for anything other than violations of consumer law, such as violations of trademark law. However, it is notable that the Court of Justice recently took this lay purchaser's perspective into account in the Louboutin v. Amazon case,[47] and didn't rule out Amazon's liability for selling counterfeit Louboutin shoes. The court specified that the operator of a marketplace can be considered to be using an identical trademark sign if the average consumer might think that they are marketing infringing goods in their own name and on their own account.

[47] Joined Cases C-148/21 and C-184/21, *Louboutin v Amazon*, CJEU, 22. December 2022; see also C-567/18, *Coty v Amazon*, CJEU, 2. April 2020, and case C-324/09 *L'Oreal v eBay*, CJEU 12. July 2011.

c) Crisis Response Mechanism

The Crisis Response Mechanism is a provision that allows the European Commission to require the largest platforms to comply with its instructions in case of a crisis. A crisis is defined by Article 36 as extraordinary circumstances that lead to a serious threat to public security or public health in the EU, such as war or pandemics. This mechanism was added to the DSA during its negotiation phase, after the EU suspended the broadcasting activities of five Russian state-owned outlets for spreading propaganda and disinformation about Russian aggression against Ukraine.[48]

Under this mechanism, the European Board of Digital Services may recommend that the Commission prompts VLOPs and VLOSEs to take specific actions when a crisis occurs. As provided by Article 36, the Commission may require that VLOPs and VLOSEs assess the possible significant contribution of their service to a serious threat or apply measures to mitigate the risk of a serious threat, such as removing war propaganda. The platforms must report their assessment, the measures taken, and any related issues to the Commission. The crisis mechanism automatically expires three months after it is triggered, unless it is renewed. The Commission is obligated to report to the Parliament and the Council on the actions taken under this mechanism.

5. Implementation of the DSA

In certain respects, the implementation of the DSA is less complicated than that of the E-Commerce Directive, as the Regulation is directly applicable and does not require a transposition law to be adopted by each Member State. However, implementing the DSA requires determining which authorities are competent to enforce it and which measures these authorities can take.

a) Competent Authorities

The competent authorities to control the implementation of the DSA are the national authorities. Member States will designate "National Coordinators of Digital Services", as per Article 49. These coordinators will receive complaints from users, have investigative powers, and may impose sanctions. They will also convene in a European Board for Digital Services, an advisory body designed to promote coordination and cooperation between them. The DSA establishes that the Member State in which an intermediary service provider is established has exclusive jurisdiction over that provider, as

[48] Council Regulation (EU) 2022/350 of 1 March 2022 amending Regulation (EU) No 833/2014 concerning restrictive measures in view of Russia's actions destabilising the situation in Ukraine; Council Decision (CFSP) 2022/351 of 1 March 2022 amending Decision 2014/512/CFSP concerning restrictive measures in view of Russia's actions destabilising the situation in Ukraine.

stated by Article 56(1). This principle aligns with the country-of-origin rule established in Article 3 of the E-Commerce Directive. Recital 123 of the DSA defines "main establishment" as the location of a provider's head or registered office, where the primary financial functions and operational control take place.

Despite these provisions, the complexity of the system established by the DSA may make it challenging to identify the authorities thar bear responsibility for enforcing it. For VLOPs and VLOSEs, the European Commission has the power to enforce the DSA in collaboration with national authorities and coordinators. The Commission holds exclusive authority over obligations that apply solely to VLOPs and VLOSEs, as outlined in Article 56(2) of the DSA. Both National Coordinators and the Commission have jurisdiction over all other DSA obligations for VLOPs and VLOSEs, as per Article 56(3). The Commission will work with national coordinators to investigate potential violations and determine whether to impose fines.

b) Enforcement

The potential severity of penalties for non-compliance with the DSA should incentivize companies to comply with its provisions. Penalties will be defined in national law and must be "effective, proportionate, and dissuasive", as provided by Article 52. The maximum penalty cannot exceed 6% of the company's annual turnover. For some infringements, such as providing incorrect or incomplete information or refusing to comply with inspections, the maximum penalty may be raised to 1% of annual turnover. Companies may also face periodic fines, with each monthly payment not exceeding 5% of their daily turnover. The European Commission can also impose fines up to these amounts. Repeated non-compliance may result in access restrictions but not a definitive ban within the EU. Under Article 82 of the DSA, the European Commission can request a regulator to ask a judicial authority to temporarily restrict user access if there is a serious and persistent breach causing significant harm and involving a criminal offense threatening people's safety or lives. Therefore, non-compliance may result in fines and temporary access restrictions at worst.

In addition to potentially high penalties for non-compliance, the DSA includes mechanisms designed to encourage compliance by the largest entities. VLOPs and VLOSEs must appoint a compliance officer, as per Article 41. VLOPs and VLOSEs must also conduct annual independent audits to assess their compliance with the DSA and submit reports to competent authorities, as outlined in Article 37. These reports must include information on measures taken to prevent the dissemination of illegal content on the platform, as well as details of the platform's internal complaint-handling system and content moderation procedures. Compliance with the provisions for controlling systemic risks by VLOPs and VLOSEs is based on an "enhanced supervision system" provided by Article 75. Under this system, the Commission may request that

very large platforms provide regulators with an action plan to address potential violations of the DSA. This action plan may include conducting an independent audit. The Commission has the discretion to determine whether the action plan is sufficient and may reject it if deemed inadequate.

Finally, the DSA stipulates in Article 45 that regulators must encourage and facilitate the development of voluntary codes of conduct at the Union level to support the proper application and ensure consistent implementation of the Regulation, particularly in regard to combating illegal content and mitigating systemic risks. The Commission and the European Board of Digital Services are tasked with promoting and supporting the creation of these codes of conduct.

III. Challenges of DSA's Implementation

The DSA is a highly ambitious piece of legislation that introduces innovative provisions aimed at addressing challenges that have not been addressed by any previous legislation. However, it is important to recognize that its implementation will undoubtedly present significant difficulties and challenges.

1. Enforcement Challenges

The DSA's highly ambitious scope and the numerous obligations it imposes, posing significant and onerous constraints for platforms,[49] bring into question the feasibility of effectively implementing this regulation and achieving its intended outcomes. While several potential challenges can be highlighted, it is important to note that this overview does not exclude the possibility of other obstacles arising.

a) Challenges in Articulating Authorities' Competencies and National Laws

The DSA framework grants the EU Commission significant supervisory and enforcement powers, a departure from the usual jurisdiction of Member States. This may be attributed to criticisms of the country-of-origin principle, which gives exclusive jurisdiction to Irish regulators since many large technology companies are based in Ireland.[50]

[49] Keller, 'The DSA's Industrial Model for Content Moderation', fn. 22, https://verfassungsblog.de/dsa-industrial-model/.

[50] Murgia, Espinoza, 'Ireland is 'Worst Bottleneck' for Enforcing EU Data Privacy Law', Irish Times, 13. September 2021, https://www.irishtimes.com/business/technology/ireland-is-%20worst-bottleneck-for-enforcing-eu-data-privacy-law-iccl-1.4672480.

The principle has actually raised controversy in the implementation of GDPR,[51] leading to the safer option of giving the Commission substantial authority over VLOPs and VLOSEs.

As a result, enforcing the DSA will necessitate coordination and collaboration between national Coordinators, the European Board of Digital Services, and the Commission. Several provisions within the DSA are specifically designed to facilitate this cooperation, such as rules that enable Coordinators and the Board to request that competent national authorities take action against a specific provider of intermediary services or that allow multiple Coordinators to conduct joint investigations, as outlined in Articles 58 and 60. While the DSA provides a framework for this necessary collaboration, it will not be a simple process and may require significant time and effort to function effectively.

Within Member states themselves, the implementation of the DSA will not be easy, as it will require the designated Digital Services Coordinator to be able to collaborate with other national authorities that may have a say on specific issues, such as the data protection authority, the competition authority, the telecommunications regulator. or even the authority in charge of enforcing consumer law. Within this framework, the effective enforcement of the DSA by national authorities appears to be delicate, considering the complexity of the task and the diverse range of digital services and platforms it covers.

The combination of the DSA with national laws may also present challenges. The DSA explicitly prohibits member states from implementing additional requirements related to its scope. As a result, it is highly likely that Germany will have to revise its NetzDG legislation, which mandates the removal of "clearly illegal" content within 24 hours after receiving a user complaint and imposes significant fines of up to 50 million euros for non-compliance.[52] However, state laws defining illegal speech will remain in effect since the DSA does not provide a definition for illegal content and defers to national laws for this determination. As a result, when implementing the DSA, providers must take into account specific provisions in relevant national laws. This requirement to adapt to the diversity of national laws can be challenging for major platforms, which have always had difficulty navigating such differences.[53]

[51] See Lancieri, 'Narrowing Data Protection's Enforcement Gap', 74 Maine Law Review, 15, 2022, https://digitalcommons.mainelaw.maine.edu/mlr/vol74/iss1/3.

[52] Griffin, 'New school speech regulation as a regulatory strategy against hate speech on social media: The case of Germany's NetzDG', 2022, 46 (9), Telecommunications Policy, 102411, https://doi.org/10.1016/j.telpol.2022.102411.

[53] Tworek, 'History explains why global content moderation cannot work', Brookings, 10. December 2021, https://www.brookings.edu/techstream/history-explains-why-global-content-moderation-cannot-work/.

It is important to acknowledge the significant resources that will be necessary for effective collaboration between regulators and for the precise monitoring of compliance with the obligations outlined in the DSA. Enforcing the DSA will come with significant costs, including the need for qualified personnel at all levels. The funding for this oversight at the EU level will come from VLOPs and VLOSEs, who will be required to pay a supervisory fee to the Commission, up to 0.05% of their annual global turnover, as specified in Article 43 of the DSA. Nevertheless, it is important to emphasize that allocating the necessary resources for the DSA to be effectively enforced will be a challenging undertaking that should not be underestimated.

b) Obstacles Arising from the Automated and Large-Scale Operations of Online Platforms

The DSA represents a significant legislative milestone as it considers the moderation practices and fundamental characteristics of today's online platforms. Such practices are primarily governed by general terms and conditions rather than legal frameworks. They are also predominantly automated through algorithmic recommendation systems, especially since the largest platforms process billions of publications daily.[54] The DSA is therefore a particularly commendable attempt to take into account these particularities of contemporary platforms in order to provide a framework that ensures both healthy online environments and respect for freedom of expression.

The transparency requirements the DSA entails are particularly important because currently available data do not provide an accurate representation of platforms' moderation practices.[55] For instance, the existing data do not shed light on potential biases that may impact platform practices. Platforms are often accused of implementing sexist or racist biases. This is partly due to the models being trained on moderation decisions made by human moderators, who are under pressure to make rapid decisions with limited time to justify them or establish facts, and can be influenced by more or less conscious prejudices.[56] There is currently a lack of clarity regarding which users are most

[54] Douek, 'Content Moderation as Systems Thinking', fn. 34, https://harvardlawreview.org/2022/12/content-moderation-as-systems-thinking/.

[55] DiResta, Edelson, Nyhan, Zuckerman, 'It's Time to Open the Black Box of Social Media', 28. April 2022; Persily, Testimony before the US Senate Committee on the Judiciary Subcommittee on Privacy, Technology and the Law, 'Platform Transparency: Understanding the Impact of Social Media', May 2, 2022; Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems', 2020, 11 (2), European Journal of Law and Technology, https://ejlt.org/index.php/ejlt/article/view/786.

[56] Griffin, 'The Sanitised Platform', 2022, 13 (1), *JIPITEC* 36, https://www.jipitec.eu/issues/jipitec-13-1-2022/5514.

susceptible to online hate speech or harassment and whether particular groups are targeted more frequently than others. Additionally, it is not clear what proportion of content is taken down without valid reason and if marginalized populations are more subject to censorship than others.[57] What is known at this point is that platforms have a primarily commercial approach to moderation and shape their moderation policies to attract the largest number of users and to ensure that users are not driven offline by objectionable or offensive content. This lack of clarity underscores the importance of the DSA's transparency requirements, which may provide much-needed insight into these issues. Although the data disclosed in the periodic reports published by platforms is limited to certain types of information, this information is intended to be supplemented by research produced by researchers who have access to the data of VLOPs and VLOSEs.

An encouraging aspect of the DSA is its recognition that content moderation systems based on machine learning models can make errors. These models operate on a probabilistic basis by predicting the likelihood that a piece of content belongs to a predetermined category. These tools may also produce false positives or false negatives. The inclusion of provisions in the DSA that require the disclosure and reduction of error rates for automated tools is thus a positive development. However, it should be noted that Article 15(1)(e) only mandates the disclosure of "indicators of the accuracy and the possible rate of error" of the automated tools used, which lacks specificity. Recital 26 of the DSA also stipulates that platforms must "take reasonable measures to ensure that the technology is sufficiently reliable to limit to the maximum extent possible the rate of errors," which implies that they must make an effort to assess the errors made by their algorithmic systems. To that end, it is crucial to establish a clear definition of "error" and a reliable methodology for evaluating error rates. For instance, let's imagine that Facebook's system removes a post featuring Courbet's painting "Origin of the World".[58] This could be considered an error because artistic works are permitted to be shared online. However, the system is not mistaken in identifying the post as featuring nudity. It could probably be agreed that errors can arise when the system flags content as violating applicable laws or terms of use when it is actually permissible, or conversely, when it fails to identify content that does violate such rules. Therefore, service providers should conduct regular tests to measure the proportion of false positives and

[57] Lux and Miss Hot Mess, 'Facebook's Hate Speech Policies Censor Marginalized Users', Wired, 14. August 2017, https://www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/.

[58] A few years ago, a French user initiated a lawsuit against Facebook alleging that the abrupt suspension of his account was due to the publication of Gustave Courbet's famous painting depicting female genitalia. However, Facebook has consistently maintained that this artwork was permissible on its platform and denied that the publication was the reason for the account's suspension. The case was finally settled. Signoret,'Censure de "L'origine du Monde": une faute de Facebook reconnue, mais pas sur le fond', Le Monde, 15 mars 2018.

false negatives, both globally and for specific content categories or user groups. The resulting information should be available to regulators, experts, and the general public to evaluate the effectiveness of algorithmic moderation and its evolution over time. In sum, while the DSA's current provisions are a welcome step, more precise measures are needed to ensure effective disclosure and reduction of error rates for automated tools.

Furthermore, despite these encouraging aspects, it is unclear if the DSA's provisions are suitable for algorithm-governed architectures that manage millions of publications posted every day by users. The practical constraints of large-scale content moderation may hinder the implementation of the DSA's procedural safeguards. Specifically, complying with Article 17's "statement of reasons" requirement for users impacted by a moderation decision may be challenging. The "statement of reasons" must clearly present the measure being taken, facts and circumstances, legal or contractual violation justifying the decision, any third-party notification, how the decision was made, and appeal options. Reasons must be provided not only for traditional moderation decisions such as account suspension, content removal, demonetization, and tagging, but also in case of demotion, downgrading, and restriction of visibility. In this respect, the scope of the rule is particularly broad, since visibility restrictions include any downgrade in ranking or recommendation systems, limitation of accessibility by other users, or "shadow banning", as specified by Recital 55. However, it is not easy to demonstrate "restriction" of visibility when the expected visibility of content is unknown.[59] Algorithmic content curation systems rely on complex scoring methods, making it hard to determine the extent of downgrading and how content would have performed without it. Within this framework, it can be difficult to ascertain the appropriate instances when users are entitled to receive accurate explanations.[60] Moreover, even when it is evident that an explanation is necessary, the complexity of machine learning models creates a substantial barrier to complete transparency.[61] Providing detailed explanations can be demanding[62] and users may not be interested in complex or complicated explanations, just as they tend to disregard Terms and Conditions. Some authors have expressed the

[59] This challenge was highlighted by several authors: Leerssen, 'An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation', fn. 35, https://www.sciencedirect.com/science/article/pii/S0267364923000018; Griffin, 'Rethinking Rights in Social Media Governance', fn. 32; Gillespie, 'Do not recommend? Reduction as a form of content moderation', Soc Media+ Soc, 2022, 8 (3), https://doi.org/10.1177/20563051221117552.

[60] Leerssen has noticed that notifying users about any decision affecting content visibility seems to prohibit methods practiced without their knowledge, such as shadow banning, Leerssen, 'An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation', fn. 35, https://www.sciencedirect.com/science/article/pii/S0267364923000018.

[61] Gillespie, 'Content moderation, AI, and the question of scale', Big Data & Society, 2020, 7, https://doi.org/10.1177/2053951720943234.

[62] Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems', fn. 56, https://ejlt.org/index.php/ejlt/article/view/786.

concern that individuals who are privileged, educated, or experienced with digital tools are more likely to exercise their rights than those who are not.[63]

On a broader level, complying with the DSA's procedural requirement will be arduous and expensive for platforms that handle millions of publications daily through automated systems.[64] These platforms will need to develop appropriate tools to handle user notifications within a reasonable timeframe, scrutinize disputed content, and notify users of any content limitations. The "statement of reasons" requirement will result in millions of moderation decisions being justified every day. Additionally, even if a small percentage of users appeal moderation decisions, there will be a need for a significant investment in human resources to handle these appeals, as they must be processed under human supervision. At the moment, platforms employ poorly paid moderators, working under intense pressure and often lacking the training necessary to understand the language and culturally relevant context in which content is published.[65] In this context, the technical and human resources that will be required for compliance with the DSA will be substantial, which raises questions as to whether platforms will be willing to make such investments.[66] Above all, there is a risk that platforms, in an effort to limit their costs, will make investment choices that are ultimately to the detriment of the objectives pursued by the DSA. Hosting providers might be tempted to invest less in controlling systemic risks, in the quality of their moderation systems, or in the fight against algorithmic biases.[67]

All in all, although the DSA's call for transparency regarding recommendation systems' parameters is a favorable development, as it can help users comprehend content curation and moderation better, it remains uncertain whether the DSA will effectively result in the re-empowerment of users, even with exceptional assurances that their fundamental rights will be safeguarded.

c) Obstacles to Accessing VLOPs' and VLOSEs' Data

The DSA allows regulators and accredited researchers to access data held by the largest platforms and search engines. Undoubtedly, this improvement will play a crucial role

[63] Griffin, 'Rethinking Rights in Social Media Governance', fn. 32; Husovec, 'Will the DSA work?', Verfassungsblog, 9. November 2022, https://verfassungsblog.de/dsa-money-effort/.

[64] Gorwa, Binns, Katzenbach, 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', 2020, 7 (1), Big Data & Society, https://doi.org/10.1177/2053951719897945.

[65] Ahmad, Greb, 'Automating social media content moderation: implications for governance and labour discretion', Work in the Global Economy, 2022, 2 (2), 176-198, https://bristoluniversitypressdigital.com/view/journals/wge/2/2/article-p176.xml.

[66] Keller, 'The EU's new Digital Services Act and the Rest of the World', fn. 22, https://verfassungsblog.de/dsa-rest-of-world.

[67] Griffin, 'Rethinking Rights in Social Media Governance', fn. 32.

in facilitating new research and enhancing comprehension of content moderation. However, it can be assumed that technology companies may be reluctant to open their black boxes.[68] They may mobilize, to oppose a too large communication of their data, rather effective legal arguments such as the need to ensure the protection of their users' personal data or the protection of their confidential information as trade secrets.

Platforms have cited the General Data Protection Regulation (Regulation (EU) 2016/679) as a barrier to sharing data with independent researchers.[69] Complying with GDPR would indeed necessitate costly legal and technical arrangements. For example, Recital 98 of the DSA states that providers should anonymize or pseudonymize personal data, except when doing so would render impossible the research purpose pursued. To tackle this challenge, a working group composed of representatives from academia, platforms, and civil society has developed a Code of Conduct, published in May 2022, which outlines how platforms can provide data and the measures researchers must take to safeguard it under the GDPR.[70] Nonetheless, the safeguarding of personal data is not the sole legal argument that platforms can use to resist the disclosure of their data. They can also claim that the requested data are confidential information protected as trade secrets, which is widely accepted for algorithms.[71]

Regulators can access data "to monitor and assess compliance" with the DSA, as provided by Article 40(1). Providers must explain to regulators "the design, logic, functioning, and the testing of their algorithmic systems, including their recommender systems" (Article 40(3)). Recital 96 adds that regulators may require access to "specific data, including data related to algorithms".[72] Although the requested data, particularly those related to algorithms, may be considered trade secrets, regulators are authorized

[68] Leerssen, 'Platform research access in Article 31 of the Digital Services Act', Verfassungsblog, 7. September 2021, https://verfassungsblog.de/power-dsa-dma-14/.

[69] Vermeulen, 'Researcher Access to Platform Data: European Developments', 2022, 1 (4), Journal of Online Trust & Safety, https://doi.org/10.54501/jots.v1i4.84.

[70] See European Digital Media Observatory, Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access, 31. May 2022, https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf.

[71] Huseinzade, 'Algorithm Transparency: How to Eat the Cake and Have It Too', European Law Blog, 27. January 2021; Maggiolino, 'EU Trade Secrets Law and Algorithmic Transparency', Bocconi Legal Studies Research Paper n°3363178, 31. March 2019; Foss-Solbrekk, 'The good, the bad and the ugly', Journal of Intellectual Property Law & Practice, 2021, vol. 16, n°3.

[72] Which includes 'the data necessary to assess the risks and possible harms brought about by the platform's or search engine's systems, data on the accuracy, functioning and testing of algorithmic systems for content moderation, recommender systems or advertising systems, including, where appropriate training data and algorithms, or data on processes and outputs of content moderation or of internal complaint-handling systems'.

to obtain access to them. As per Article 1(2)(b) of the EU Trade Secrets Directive,[73] the protection of trade secrets does not override regulations requiring trade secret holders to reveal information to authorities to carry out their duties. The EU Commission or Digital Services Coordinators will thus be able to proceed as in the Google Shopping case[74], where the Commission examined Google's model and found it was designed to favor Google Shopping over other comparison sites.

The situation for researchers is different. The DSA stipulates that VLOPs and VLOSEs must provide access to data to vetted researchers "upon a reasoned request from the Digital Services Coordinator". This access is for "the sole purpose of conducting research that contributes to the detection, identification, and understanding of systemic risks (...) and to the assessment of the adequacy, efficiency, and impact of the risk mitigation measures" (Article 40(4)). Providers facing an access request for researchers can request the Digital Services Coordinator to amend the request if they do not have access to the data or if "giving access to the data will lead to significant vulnerabilities in the security of their service or the protection of confidential information, in particular trade secrets" (Article 40 (5)). The Digital Services Coordinator decides on the provider's application to amend the access request, provided that this application contains proposals for alternative means of providing access to the requested data or other appropriate data (art. 40 (6)).

Within this framework, it is conceivable that large platforms that do not wish to open up access to their data too widely may argue that certain information is not necessary to assess their systemic risks or mitigation measures. But they may also be tempted to invoke the protection of trade secrets. Indeed, the definition of trade secrets is particularly broad. Under Article 2 of the EU Trade Secrets Directive, trade secrets are pieces of information that are secret, have commercial value due to their secrecy, and are subject to reasonable steps to keep them secret. A lot of data can correspond to this definition. It is also accepted that algorithms can be protected as trade secrets.[75] It is, moreover, generally considered that the training data and any other information relating to the algorithm falls within the protection[76]. It is very likely that researchers wanting to assess systemic risks will need this type of information and will be confronted with the protection of trade secrets.

[73] Directive (EU) 2016/943 of the European Parliament and of the Council of 8. June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use, and disclosure (OJ L 157, 15. June 2016, p. 1), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0943.

[74] EU Commission Decision, 2017, Case AT.39740, *Google Search (Shopping)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1516198535804&uri=CELEX:52018XC0112(01).

[75] Maggiolino, 'EU Trade Secrets Law and Algorithmic Transparency', fn 72.

[76] Ibid.

Researchers, even vetted ones, are not considered "administrative or judicial authorities" performing their duties under Article 1(2)(b) of the Trade Secrets Directive. Yet Article 5 of this Directive allows disclosure of trade secrets to protect "a legitimate interest recognised by Union or national law". However, the DSA repeatedly emphasizes the need to protect trade secrets. Recital 97 states that access to data should protect "the rights and legitimate interests" of VLOPs and VLOSEs and their users, including "personal data, trade secrets, and other confidential information". While the DSA provides that consideration of the commercial interests of providers should not lead to a refusal to provide access to data, this is imposed "without prejudice to Directive (EU) 2016/943 of the European Parliament and of the Council" (the EU Trade Secrets Directive). In other words, the DSA does not in any way state that the objective pursued by vetted researchers constitutes an interest that must take precedence over the protection of trade secrets. On the contrary, it seems to reserve, in a formulation that can be found vague, the safeguarding of trade secrets.

The DSA itself appears to anticipate that clarification is needed and specifies that elaboration will be provided on this point. Article 40(13) provides that the Commission will, after consulting the Board, adopt delegated acts "laying down the technical conditions under which providers (...) are to share data". Those delegated acts will lay down "the specific conditions under which such sharing of data with researchers can take place" in compliance with GDPR, "as well as relevant objective indicators, procedures, and, where necessary, independent advisory mechanisms in support of sharing of data, taking into account the rights and interests of the providers (..) and the recipients of the service concerned, including the protection of confidential information, in particular trade secrets, and maintaining the security of their service". It is therefore to be hoped that these delegated acts will clarify and guarantee researchers' access to the data they need.

One might find hope in the fact that the largest platforms (Google Search, YouTube, Twitter, TikTok, Microsoft Bing, Linkedin, Meta, and Instagram) have already committed to sharing data with researchers under the EU Code of Practice on Disinformation.[77] In particular, they have pledged to support research on misinformation and to refrain from taking action against researchers, users, or accounts that undertake or participate in this research. The signatories to the code have further committed to developing, funding, and cooperating with an independent third-party body that can monitor researchers and research proposals. The DSA also mentions this possibility, as outlined by Article 40(13), which permits delegated acts to establish independent advisory mechanisms to facilitate the sharing of data. However, presently, this

[77] Vermeulen, 'Researcher Access to Platform Data: European Developments', fn. 70, https://doi.org/10.54501/jots.v1i4.84.

mechanism can only be considered for the specific purpose of conducting research on misinformation.

2. Issues Raised by the Measures to Address Systemic Risks

The DSA introduces a new regime for managing risks by tackling issues at a systemic level and fostering dialogue between regulators and platforms. Under this regime, VLOPs and VLOSEs are mandated to furnish regulators with their evaluations of systemic risks whenever requested. Additionally, regulators will be given access to research conducted by vetted researchers who have been granted access to platform data. Regulators will publish annual assessments of systemic risks, guidelines, best practices, and recommended actions.

It's worth emphasizing that the DSA represents a milestone in the effort to tackle not only content that is deemed illegal according to national laws but also material that might not be illegal but poses undesirable risks. However, the definition of systemic risks in Article 34(1) is particularly broad and vague. For instance, anything that has the potential to undermine fundamental rights could be categorized as a systemic risk, implying that any type of content could eventually be linked to such a risk. This creates significant uncertainty about what platforms, researchers, and regulators might classify as systemic risks. A comprehensive interpretation of "systemic risks" could potentially lead to the emergence of a "grey" category of content that is not explicitly illegal but deemed undesirable due to its potential to create systemic risks. Furthermore, the creation of this "grey" category, halfway between the illegal and the legal, presents a real threat to freedom of expression. Could platforms be incited to censor content excessively to minimize systemic risks? It appears that Articles 34 and 35 could lead to greater content removal rather than a reduction in unjustified censorship.

There is also considerable uncertainty as to how these provisions will be implemented. Although the DSA includes a vast array of mitigation measures, it does not specify which ones should take precedence and affords platforms substantial leeway. The role of European and national regulators in this regard is not entirely clear, as it only entails publishing reports containing guidelines, best practices, and recommendations, but nothing that is notably binding. Of course, platforms may be compelled to devise an "action plan" under the "enhanced supervision system" outlined in Article 75, but it is challenging to determine at this stage what such an "action plan" should cover.

Of course, the actual application of these provisions will play a pivotal role in determining their efficacy. The Commission will play a decisive role in this respect, although its role is primarily limited to collaborating with the European Board of Digital Services to produce annual reports outlining best practices for mitigating systemic risks and issuing guidelines and recommended actions in partnership with Digital Services Coordinators. An interesting question is whether regulators will take a stance on platforms'

terms of use or the substance of online content policies. At this stage, it is possible to say that the Commission may promote the development of codes of conduct, as outlined in Article 45. However, it remains uncertain whether regulators will be willing to offer specific recommendations and provide close guidance to platforms. All in all, it is disappointing that the provisions on controlling systemic risks and fostering dialogue between platforms and regulators on this matter are not more precise, and there is limited guidance on how this dialogue should take place.

3. Challenges of Applying the DSA to Emerging Technologies

The DSA is an innovative piece of legislation that addresses the unique characteristics of modern platforms. It recognizes the role of private standards in regulating speech at a global scale and the widespread use of algorithmic systems in content curation and moderation. The DSA also introduces a new regulatory approach based on information exchange and dialogue between regulators and companies, as well as the involvement of independent third parties such as trusted flaggers, vetted researchers, and out-of-court settlement bodies. However, there are concerns about whether the DSA's focus on major platforms and search engines makes it suitable for emerging technologies and new online environments.

It is worth considering whether the DSA will effectively address the challenges posed by extended reality, which encompasses both virtual reality and augmented reality.[78] Of course, the DSA has the advantage of recognizing the significance of terms and conditions in shaping users' experiences and content policies. In this regard, it foresees the need for contractual agreements to play a more critical role than statutes in regulating the operations and use of virtual reality environments. Indeed, the DSA mandates platforms to "codify" their content moderation rules in precise and unequivocal language under Article 14. This provides a crucial assurance to users of immersive environments, where these private regulations will have a more significant impact. Additionally, the DSA's systemic risk management approach should be particularly beneficial in dealing with the novel threats posed by highly immersive technologies, which enable the detection and tracking of users' slightest movements, including facial expressions, and increase the risks of manipulation.

However, the procedural rules laid out by the DSA may not be entirely suited to immersive environments, where communication between users occurs through non-verbal or oral means. Moreover, the actions and interactions of users will probably not be recorded, which renders any principles on content removal or visibility restrictions

[78] Lemley, Volokh, Law, 'Virtual Reality, and Augmented Reality', 166 University of Pennsylvania Law Review, 1051, 2018, https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9622.

impractical. Furthermore, identifying and penalizing inappropriate conduct will require the development of new mechanisms not accounted for in current regulations. For instance, Meta has added a "personal boundary" system[79] to its Horizon virtual reality experiences, aiming to stop harassment in Virtual Reality. Yet it will probably also be necessary to explore real-time alert systems to prevent immoral or unlawful conduct. Additionally, specific conflict resolution methods will need to be devised since the out-of-court settlement system provided for by the DSA lacks provisions for binding decisions.

Furthermore, the DSA is designed to apply mainly to the large centralized technology companies that dominate today's Internet. Many of its obligations do not apply to small businesses. The possibility of developing distributed or decentralized networks does not seem to be really envisaged by the Regulation. Of course, the DSA expressly provides that users of online platforms must be able to choose the recommendation system that suits them and that VLOPs and VLOSEs must always offer them a recommendation system that does not rely on profiling. These rules allow users some degree of choice but these choices are still made within centralized platforms that the DSA aims to regulate. They do not address the challenge of regulating decentralized or distributed social networks that will hardly be subject to the DSA's rules. It is certainly understandable that regulators wanted to avoid placing too many restrictions on small entities.[80] However it will, sooner or later, be necessary to establish a clear legal framework for moderation practices on distributed or decentralized networks.[81]

Overall, the DSA represents a new and pragmatic regulatory approach that acknowledges the challenges posed by contemporary platforms. It strikes a delicate balance between combatting harmful content and safeguarding freedom of expression. The DSA acknowledges the crucial role of platform operators who have the technical expertise and control over the platform architecture to effectively enforce online regulations. It provides a preliminary framework for how platforms can manage speech and behavior online through their contractual terms and conditions. Despite imposing significant restrictions, particularly on larger platforms, the DSA fosters a constructive dialogue between platforms and regulators. Critics may argue that the DSA places too much emphasis on procedural guarantees that may prove difficult to implement in practice, and that its application may lead to uncertainties. There is no doubt that implementing and applying the DSA will inevitably present challenges. Nonetheless, it is indisputable that the DSA represents a significant and unprecedented step forward in

[79] https://www.theverge.com/2022/2/4/22917722/meta-horizon-worlds-venues-metaverse-harassment-groping-personal-boundary-feature.

[80] Komaitis, de Franssu, 'Can Mastodon Survive Europe's Digital Services Act?', Tech Policy Press, 16. November 2022, https://techpolicy.press/can-mastodon-survive-europes-digital-services-act/.

[81] Rozenshtein, 'Moderating the Fediverse: Content Moderation on Distributed Social Media', 2 Journal of Free Speech Law, 2023, Forthcoming, https://ssrn.com/abstract=4213674.

regulating online platforms, even though its effectiveness in enhancing online behavior and platform practices over the long term remains to be seen.

Impacts of the Digital Services Act on the Facebook "Hate Speech" Decision by the German Federal Court of Justice

Ruth Janal

I. Introduction

In July of 2021, the German Federal Court of Justice (Bundesgerichtshof – BGH) rendered two decisions regarding Facebook's right to ban and delete hate speech from their platform. Since party names remain anonymous when German decisions are published, the Facebook decision has been dubbed the "Hate Speech" decision. In a nutshell, the BGH stated that providers of social network services are entitled to set their own communication standards and to restrict legal forms or speech. However, the service provider's standard terms and conditions must set out the network's communication standard in a transparent, factual, and non-arbitrary way and must provide procedural safeguards for users whose speech is being restricted by means of content moderation. Meanwhile, the EU's Digital Services Act (DSA)[1] also includes provisions on standard terms and conditions and provides for complaints mechanisms against content moderation decisions.

In the following article, I will explore how the BGH case law and the provisions of the DSA interact and will argue that the BGH has to adapt its case law once the DSA enters into operation on 17 February 2024. I start by giving a brief explanation of the BGH's Facebook decisions (II), then turn to the corresponding rules of the DSA (III), and finally discuss the DSA's implications for the Court's case law (IV).

II. The BGH's "Hate Speech" Decision

1. The Right to Restrict Lawful but Awful Content

Basically, all online platforms restrict so-called "lawful but awful" speech. The impact of the Ukraine war on social media illustrates why it can be desirable for intermediary

[1] Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market For Digital Services and amending Directive 2000/31/EC, Digital Services Act, OJ L, 2022, 277/1.

service providers (ISPs) to restrict the use of their services and exclude various forms of lawful speech. Many such posts, while lawful, do not appeal to a majority of the provider's users, may render the service unattractive to advertising partners, and may make the provider subject to political and societal pressure. This includes depictions of violence and death, disinformation by war parties and information on the war that might put soldiers in harm's way, as well as gender-based hate speech towards politicians, experts, activists, and journalists alike. In its "Hate Speech" decision, the BGH acknowledges and accepts Facebook's interests in limiting certain kinds of lawful speech. The reason is two-fold: First, despite its considerable economic power and social influence, Facebook is not a state actor and thus not directly bound by the fundamental rights provided for under the German constitution (the "Basic Law"). Secondly, Facebook as an economic entity is itself entitled to rely upon the freedom to conduct a business flowing from Art. 12 of the German Basic Law.

2. Horizontal Effects of Fundamental Rights

Modern fundamental rights doctrine accepts that fundamental rights not only serve as protections against state actors but may also have an indirect horizontal effect on relations between private parties.[2] In German civil law, provisions on good faith in contractual relations (§§ 138, 242, 826 of the BGB[3]) and on the control of standard terms and conditions (§ 307 BGB) may serve as points of entry for the consideration of fundamental rights in private relations. In the "Hate Speech" decision, the BGH chose to rely upon the rules for a judicial review of unfair contract terms (§ 307 BGB) as an instrument to require Facebook to undertake a balancing of fundamental rights.

Under German doctrine, the extent of a horizontal effect of fundamental rights depends upon a variety of factors, most notably a structural imbalance between the parties, the social power of one of the actors, the significance of the service concerned, and the inevitability of the situation.[4] In the run up to the "Hate Speech" decision, some lower courts and various academics had argued that Facebook should not be allowed to restrict any lawful speech on its platform. Rather, it was claimed, Facebook should be required to respect freedom of speech in the same way as a state actor would, due to the social network's pre-eminent position for the distribution of news and the forming of

[2] See for a discussion on the indirect horizontal effects of fundamental rights on ISPs in the Netherlands Quintais, Appelman and Fahy 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation', 25 November 2022, https://ssrn.com/abstract=4286147, p. 22 et seq.

[3] Bürgerliches Gesetzbuch – the German Civil Code.

[4] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, para 67.

societal opinions.[5] The BGH did not concur.[6] The Court highlighted that Facebook is not the only social network on the market and that there are options to express one's opinion outside of social media. This notwithstanding, the Court found that Facebook possesses dominant market power in the market for social networks and provides a significant means of communication on the internet, leading to lock-in effects for its users.[7] Facebook must therefore consider the multipolar interests of its users (both those posting and those viewing content) as well as the interests of parties affected by the content on the platform.[8]

3. Consequences for the Design of Facebook's Standard Terms

The BGH employed the judicial scrutiny of standard terms and conditions as a lever to provide guardrails for the design of Facebook's standard terms and content moderation policies. The Court emphasized two pillars: transparency and procedure. First, the provider's standard terms and conditions must set out factual grounds for any restriction of lawful content in a transparent, non-arbitrary way. Facebook may impose a communication standard that restricts *the way* users articulate their speech, thus providing for a safe online environment. The Court seemed hesitant to declare that the platform may also restrict particular *viewpoints* (i.e., specific political content). Secondly, the platform's standard terms and conditions must stipulate procedural safeguards for the blocking of content and the suspension of accounts. The Court required Facebook to provide an internal complaints mechanism that consists of (1) a notification of the action taken, (2) a statement of reasons, (3) an opportunity for the user to reply, and (4) a review of the decision following a user's complaint. In the case of the removal of content, the notification may be given after the removal has occurred, but the content may not be permanently deleted until the user has been heard. In case of the suspension of an account (even if temporarily or partially), the user must be heard before the suspension occurs.

---

[5] OLG München, Case 18 W 1281/19, MMR 2021, 79 paras 72, 74; OLG Oldenburg, Case 13 W 16/19, 2020; MMR 41 para 9; KG Berlin, Case 10 W 172/18, MMR 2020, 47 paras 17, 19; for platforms with a dominant market position Elsaß, Labusga and Tichy 'Löschungen und Sperrungen von Beiträgen und Nutzerprofilen durch die Betreiber sozialer Netzwerke ', CR 2017, 234, 241; Specht, 'Die Entwicklung des IT-Rechts im Jahr 2018', NJW 2018, 3686, 3687.

[6] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347 para 59. See similarly Beurskens 'Hate-Speech – zwischen Löschungsrecht und Veröffentlichungspflicht', NJW 2018, 3418, 3420; Friehe 'Löschen und Sperren in sozialen Netzwerken', NJW 2020, 1697, 1702; Holznagel 'Overblocking durch User Generated Content (UGC)-Plattformen: Ansprüche der Nutzer auf Wiederherstellung oder Schadensersatz?', CR 2018, 369 para 20; Spindler 'Löschung und Sperrung von Inhalten aufgrund von Teilnahmebedingungen sozialer Netzwerke', CR 2019, 238 paras 23, 33, 35.

[7] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, paras 71, 77 et seq.

[8] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, paras 59 et seq.

Thus, even though the German Civil Code does not contain provisions for content moderation practices, the BGH leveraged the judicial scrutiny of standard terms and conditions to put into place procedural safeguards for freedom of speech on social media platforms. Since Facebook's standard terms did not comply with these requirements, the Court deemed those terms as contrary to the requirement of good faith and thus invalid under § 307 BGB. In effect, this meant that users were able to require the restoration of blocked content.[9]

4. Open Questions

As is often the case with important case law, the "Hate Speech" decision raises almost more questions than it provides answers. I want to address four important points: the decision's implications for a) other online platforms and b) for illegal speech, c) the authority of interpretation for internal communication standards, and d) whether platforms may properly terminate their contract with individual users.

a) Implications for Other Online Platforms

The decision's implications for online platforms other than Facebook are unclear. The Court does not address situations in which other ISPs might be held to a similar standard as Facebook. In arguing that Facebook has to provide for a complaints mechanism by virtue of its standard terms, the BGH relied upon the obiter dictum that Facebook wields significant market and social power and that users may be "locked in" by their social connections within the network.[10] However, the court did not establish any criteria for determining significant market and social power. Nor did the court in any way specify the required extent of lock-in effects. The regrettable lack of specificity is evident when one considers whether ISPs such as Twitter or TikTok are also required to implement a complaints mechanism. Twitter does not possess a dominant market position, but its societal influence may exceed that of Facebook – at least in certain circles, such as in the media and science communities. TikTok on the other hand possesses a strong economic position, but its lock-in effects may be lower than Facebook's, as it thrives on entertainment, not on social connections.

The only information one may glimpse from the "Hate Speech" decision regarding ISPs other than Facebook is that the BGH will hold access providers to a higher standard than social networks, as they "provide access to the internet as such".[11] A suspension

[9] OLG Stuttgart, Case 4 U 484/20 para 112 et seq.; OLG Celle, Case 13 U 84/19, MMR 2022, 399 para 74; an opportunity to make up for procedural failings is allowed by OLG Frankfurt, Case 16 U 229/29, ZUM-RD 2022, 630, para 100 et seq.

[10] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, para 79.

[11] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, paras 67, 71.

of service by an access provider eliminates the user's option to participate in online speech *per se* and will thus face higher judicial scrutiny.

b) Blocking of Illegal Content

The "Hate Speech" decision concerned content which was undisputably legal, even though it may have been in contravention of Facebook's community standards. What are the implications of the decision for Facebook's measures sanctioning illegal content? Firstly, the transparency requirements of the "Hate Speech" decision do not translate to removal or blocking decisions on the basis of *illegal* content. Whether content is lawful or not is a question of law. There is no point in repeating a description of illegal content and/or practices in a provider's standard terms. In contrast, the procedural requirements imposed by the BGH seem also helpful when dealing with (supposedly) illegal content: In determining whether content is illegal, platforms and their employees may make mistakes. Thus, a requirement to notify users and establish a complaints mechanism seems reasonable also with respect to illegal content.

c) Correct Interpretation of the Communication Standard

Undoubtably, the correct interpretation of the law lies with the courts. Thus, the courts have the final voice in determining which content is illegal. But who possesses the authority of interpretation when it comes to a provider's communication standard? Usually, in describing unwanted speech, such standard terms contain plenty of ambiguous terms, such as "hateful" or "violent". I believe that platforms should be granted some discretion in applying their own communication standards, provided the interpretation occurs in a non-arbitrary manner.

The BGH's focus on transparency and procedural safeguards seems to imply that the Court wanted to grant Facebook some discretion in interpreting its own community standards,[12] provided these are sufficiently transparent for users. This makes a lot of sense. We shouldn't aspire for detailed legal commentary or handbooks on the "Facebook Community Standards" authored by the courts. Internal guidelines on content moderation are usually extensive[13] and are constantly being adapted, taking into con-

[12] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, para 117.

[13] Ortolani 'The resolution of content moderation disputes under the Digital Services Act', 1. February 2023, https://ssrn.com/abstract=4356598, p. 13.

sideration current harmful trends and secret communication codes that are community-specific. Thus, some discretion in applying content moderation policies seems warranted.[14]

It should be noted that giving the platforms such leeway runs counter to the general principle of interpreting standard terms *contra proferentem*: Art. 5 sent. 2 of the Unfair Terms Directive declares (for B2C-contracts) that "where there is doubt about the meaning of a term, the interpretation most favourable to the recipient of the service shall prevail".[15] However, this directive was designed with contracts for the bi-direction exchange of goods and services in mind, not for the platforms of opinions that must balance the rights and interests of the affected user, all their other users, and third parties.

d) Proper Termination of Contract

Finally, the "Hate Speech" decision does not address whether providers are allowed to properly terminate a user contract without cause. It seems implicit in the "Hate Speech" decision that the freedom to contract does not apply to social media companies with a dominant position. Otherwise, it would not have mattered whether Facebook was entitled to suspend the claimant's account on the basis of supposed hate speech, since Facebook could have simply terminated the respective user account without cause. However, the BGH does not state explicitly that proper termination of the contract is out of the question. Nor does the Court explain under which circumstances such a limitation to the freedom to contract arises. Traditionally, German law ponders such restrictions in cases that involve entities with a dominant market position where essential services are concerned.[16]

III. Corresponding Provisions in the Digital Services Act

1. Overview of Relevant DSA Provisions

As has been explained above, the BGH based its "Hate Speech" decision on a rather freewheeling judicial scrutiny of standard terms, because the German Civil Code does

[14] Furthermore, "No judicial authority could be realistically expected to sustain the decision-making frequency that social media platforms manage through their content moderation procedures", cf. Ortolani, 'The resolution of content moderation disputes under the Digital Services Act', 1. February 2023, https://ssrn.com/abstract=4356598, p. 3.

[15] Council Directive 93/13/EEC on unfair terms in consumer contracts, 1993, OJ L 95/29; the German implementation can be found in § 305c II BGB.

[16] Wagner, MüKo BGB, 8th ed. 2020, § 826 BGB paras 216 et seq.

not contain any rules regarding content moderation. On 17 February 2024, the legal landscape on content moderation will change significantly. Instead of a dearth of regulations, we will face a regulatory thicket. I say regulatory thicket, as the DSA contains many diverse puzzle pieces on content moderation and procedural rights of both users and persons affected by illegal or harmful content. Unfortunately, it is not entirely clear how these pieces are supposed to fit together. In the following, I will provide a brief overview of the relevant DSA provisions and will then delve deeper and compare those provisions to the BGH "Hate Speech" decision.

The DSA contains three provisions that correspond to the "Hate Speech" ruling: Art. 14, Art. 17, and Art. 20. Art. 14 DSA presupposes that ISPs are allowed to impose restrictions on the use of their service and set their own communication standards (on the basis of the principle of freedom of contract[17]). Speech restrictions as well as the complaints handling mechanism must be set out in a transparent and user-friendly way in the provider's standard terms and conditions. Art. 17 DSA requires that host providers must give a statement of reasons in case of a restriction of visibility of specific content (i.e., removal, disabling of access, or demotion of content). This is also true if a user's monetization opportunities are restricted or if a user account or the provision of the service is suspended or terminated. Finally, online platform services[18] must implement a complaint-handling system under Art. 20 DSA, (except for micro and small enterprises).[19] These provisions are complemented by obligations for online platforms to submit to out-of-court dispute settlements (Art. 21 DSA) and to introduce measures and protection against misuse (Art. 23 DSA). The DSA does not, however, address the issue of proper termination of service contracts without reason.

2. Commonalities and Differences with the "Hate Speech" Ruling

At first glance, the DSA seems to simply provide an update to the requirements spelled out by the BGH "Hate Speech" ruling. The horizontal effect of fundamental rights underlies both German case law and the Digital Services Act.[20] Further, both follow a

[17] Recital 45.

[18] Art. 2 (h) DSA defines an online platform as 'a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public'. As recital 13 explains, this does not include cloud computing services and webhosting services 'when serving as infrastructure'.

[19] Art. 19 (1) DSA, Art. 2 (2), (3) Commission Recommendation 2003/361/EC concerning the definition of micro, small and medium-sized enterprises, OJ L 124/36.

[20] For the DSA see Quintais, Appelman and Fahy 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation', 25. November 2022, https://ssrn.com/abstract=4286147, p. 25.

'procedure before substance' approach[21] to ensure a balancing of rights. Nonetheless, a closer look also reveals some important differences.

a) Right to Set a Communication Standard

Let us start with an important common ground. Both the "Hate Speech" ruling and the DSA operate under the same premise:[22] Providers of intermediary services benefit from the principle of freedom of contract and are thus allowed to restrict the use of their service. ISPs may therefore set their own communication or marketplace standards and restrict legal speech and/or legal activities. However, they must do so in a transparent,[23] proportionate, and non-arbitrary manner[24] and have due regard to the rights and legitimate interests of all parties involved.[25]

At first glance, it is perplexing that Art. 14 (4) DSA requires providers to act in "a diligent, objective and proportionate manner" only "in applying and enforcing" the restrictions of the service. However, recital 47 clarifies that those requirements apply also in the design process. Arguably, the design of standard terms and conditions is a matter covered by the Unfair Contract Terms Directive for consumer contracts and by Art. 3 (1) (c) of the Platform to Business Directive.[26] This could explain why Art. 14 (4) DSA does not refer to the design process.[27]

[21] Ortolani 'The resolution of content moderation disputes under the Digital Services Act', 1. February 2023, https://ssrn.com/abstract=4356598, p. 4, 7. For this 'procedural turn' in the P2B-Directive see Busch, 'The P2B Regulation (EU) 2019/1150: Towards a "procedural turn" in EU platform regulation?' EuCML 2020, 133, 134.

[22] Raue and Heesen, 'Der Digital Services Act', NJW 2022, 3537, para. 22.

[23] For a detailed analysis of Art. 14 DSA see Quintais, Appelman and Fahy, 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation', 25. November 2022, https://ssrn.com/abstract=4286147, p. 11 et seq. For a discussion of the 'impossible nature of the duties online providers are faced with' since complex information can hardly be communicated clearly and concisely see Lodder and Carvalho, 'Online platforms: Towards an information tsunami with new requirements on moderation, ranking, and traceability', 4. March 2022, https://ssrn.com/abstract=4050115.

[24] Art. 14 (1), recital 47 DSA, Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, paras 61, 81.

[25] Case III ZR 179/20 'Hate Speech', BGHZ 230, 347, paras 69, 75; Art. 14 (4) DSA. Recital 47 DSA highlights that providers of very large online platforms in particular should pay due regard to freedom of expression and of information, including media freedom and pluralism. As to the applicable human rights standards see Quintais, Appelman and Fahy 'Using Terms and Conditions to Apply Fundamental Rights to Content Moderation', 25. November 2022, https://ssrn.com/abstract=4286147, p. 17 et seq.

[26] Regulation (EU) 2019/1150 of the European Parliament and of the Council on promoting fairness and transparency for business users of online intermediation services OJ L 2019, 186/57.

[27] According to Art. 2 (4) (e) and (f) DSA, the DSA is without prejudice to the P2B-Directive and to Union law on consumer protection.

b) Addressees

A major difference between the "Hate Speech" decision and the DSA rules pertains to the circle of service providers that must balance freedom of speech considerations. According to the BGH, the judicial scrutiny of unfair contract terms will vary, depending upon the dominant market and societal position of a service provider, perceived lock-in effects, and the essentiality of the service. As a consequence, the judicial scrutiny of standard terms will be less restrictive when dealing with an insignificant social network and will be more restrictive when the service rendered is an essential service, such as access to the internet.

The application of the DSA provisions depends upon the nature and size of the service: Art. 14 requires all ISPs to explain restrictions of their service and their complaints handling mechanisms in transparent and user-friendly terms. In contrast, the requirement to provide a statement of reasons for removal, restriction, or suspension decisions only applies to host providers. This distinction is particularly curious in light of the ECJ's *UPC Telekabel* decision, in which the court held that internet users must be given the possibility to challenge blocking orders against access providers in court – which is only possible if the owner of a blocked website has been notified of the blocking decision.[28] Finally, the DSA requires specific ISPs to provide for an internet complaints mechanism and makes them subject to an out-of-court dispute settlement process. These requirements only apply to online platforms, a sub-category of host providers that consists mainly of social networks and online marketplaces (except for micro and small enterprises).[29] The distinctions drawn by the DSA between different types of service providers are neither self-explanatory nor entirely satisfactory. At a minimum, however, the DSA provides for more legal certainty than the "Hate Speech" ruling, as it does not rely upon vague criteria such as "dominant market position", "societal power", or "significance of services".

c) Notification and Statement of Reasons

With respect to procedural safeguards, the DSA is much more specific than the BGH's case law. Under Art. 17 DSA, a host provider has to notify the affected user if the provider sanctions the user for content which is illegal or incompatible with the provider's

[28] Case C-314/12, *UPC Telekabel Wien v. Constantin Film Verleih*.

[29] Art. 2 (i) DSA defines online platforms as 'online platform means a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation'. Further guidance is provided by recital 13 et seq.

terms and conditions. Whereas the "Hate Speech" ruling was concerned with the removal of content and the temporary suspension of accounts, Art. 17 DSA applies to an all-encompassing array of sanctions imposed by the host provider: any restriction of the visibility of specific content (i.e., removal, age-verification, disabling of access or demotion);[30] the suspension, termination or other restriction of monetization; suspension or termination of the provision of the service in whole or in part (i.e., a "read-only" mode of accounts); or suspension or termination of the user's account. The notification requirement does not arise if the service provider is acting upon a court order (Art. 17 (5) DSA) or if the content "is deceptive high-volume commercial content". Recital 45 explains that this exception is meant to cover the inauthentic use of the service via bots or fake accounts. Arguably, the exception may also apply to spam, but only if the content is deceptive, which is not always the case.

The notification must contain a statement of reasons that is specific and meaningful.[31] According to Art. 17 (3) DSA, the user must be informed of the nature of the measure being taken as well as the intended duration and territorial scope. Further, the statement must name the facts and circumstances relied upon in taking the decision and cite which legal or contractual provisions were supposedly violated. The use of automated means in the process must be disclosed. Finally, the user must be informed of the possibilities for redress (internal complaint-handling mechanisms as well as out-of-court dispute settlement and judicial redress). In sum, Art. 17 (3) DSA is much more specific than the guidance provided by the "Hate Speech" ruling, which solely states that the provider needs to give reasons for its decision.

Art. 17 (4) DSA adds that the statement of reasons "shall be clear and easily comprehensible and as precise and specific as reasonably possible under the given circumstances". In particular, the information must allow for an effective exercise of redress on the part of the user.

As to the timing of the notice, Art. 17 (2) stipulates that the notification must be given "at the latest from the date that the restriction is imposed". Thus, even for account suspensions or service restrictions, it is sufficient that the user is notified after the sanction has been imposed (and heard only subsequently).[32] This is a divergence from the "Hate Speech" ruling in which the BGH opined that users must be heard before any account suspension (even if the suspension is temporary or partial).

[30] Conversely, fact-checking labels do not seem to be included, cf. Holznagel, 'Zu starke Nutzerrechte in Art. 17 und 18 DSA', CR 2022, 594, 596.

[31] For an in-depth analysis see Ortolani 'The resolution of content moderation disputes under the Digital Services Act', 1. February 2023, https://ssrn.com/abstract=4356598, p. 11 et seq.

[32] Raue and Heesen, 'Der Digital Services Act', NJW 2022, 3537, para. 31.

d) Internal Complaints Handling System

Many online platforms already provide internal complaints systems that enjoy widespread popularity.[33] Under Art. 20 DSA, online platforms will be required by law to implement an internal complaints handling system. The complaints mechanism is available to two groups: (1) users who have been sanctioned by the platform and (2) parties who have notified the platform of content which violates the law or the platform's terms and conditions.[34] The complaints mechanism must be available for at least six months after the notification of the decision taken by the online platform. The mechanism relates to decisions on whether or not to

- disable access to or restrict visibility of the information;
- suspend, terminate, or otherwise restrict the ability to monetise information;
- suspend or terminate the provision of the service, in whole or in part;
- suspend or terminate the recipients' account.

The design of the complaints mechanism is left to the discretion of the online platform.[35] Art. 20 solely requires that the process be "easy to access" and "user-friendly".

Platforms must review their original decision in a timely, non-discriminatory, diligent, and non-arbitrary manner and provide the complainant with a statement of reasons and information on further redress options (out-of-court settlement, judicial redress).[36] The complaints mechanism requires a "human in the loop", as it cannot be based solely upon automated means, and the decisions must be made "under the supervision of appropriately qualified staff". Arguably, this implies that neither entry-level staff nor the automated decision making system used must be appropriately qualified, but solely appropriately supervised.

Art. 20 (4) DSA requires providers to reverse their decision without undue delay if a complaint "contains sufficient grounds for the provider of the online platform to consider that its decision not to act upon the notice is unfounded or that the information to which the complaint relates is not illegal and is not incompatible with its

[33] For an analysis of various transparency reports see Holznagel, 'Zu starke Nutzerrechte in Art. 17 und 18 DSA', CR 2022, 594, 597.

[34] According to Holznagel, 'Zu starke Nutzerrechte in Art. 17 und 18 DSA', CR 2022, 594, 596, the complaints mechanism is open to notice-givers only if they are also recipients of the service. In my view, however, Art. 20 (1) defines notice-givers as recipients of the service for the purpose of Art. 20. Cf. also Art. 3 (p) DSA.

[35] Ortolani 'The resolution of content moderation disputes under the Digital Services Act', 1. February 2023, https://ssrn.com/abstract=4356598, p. 11.

[36] Art. 20 (4), (5) DSA.

terms and conditions, or contains information indicating that the complainant's conduct does not warrant the measure taken". Art. 20 DSA fails to take into account that content moderation decisions often relate to a three-party-conflict: user, platform, and party affected by the content (and/or flagger). The review process is based upon the complaint issued and does not have to take into account the third party's position.[37]

Since Art. 54 DSA grants recipients of the service a right to compensation for any infringement by a provider of its DSA obligations,[38] it is a pressing question whether Art. 20 (4) DSA contains an obligation to render a correct review decision or merely an obligation to act diligently in the review process. The wording of Art. 20 (4) DSA seems to imply that providers must render a correct review decision. Nonetheless, a systemic analysis of the DSA shows that the review requirement is merely an obligation to reconsider the decision in a timely, non-discriminatory, diligent, and non-arbitrary manner.

It is important to read Art. 20 DSA in the context of Art. 6 DSA, so allow me a short digression on Art. 6 DSA. Keeping with the tradition of the eCommerce-Directive,[39] Art. 3 et seq. DSA provide liability shields for specific ISPs. By virtue of Art. 6 DSA, host providers are exempt from any liability as long as they do not have knowledge of the illegal content and act expeditiously to remove or disable access to the illegal content once such knowledge is obtained. The DSA does not contain an obligation for host providers to take down content! Rather, host providers run the risk of being held liable under Member States' laws if they do not remove illegal content and cannot rely upon the liability shield of Art. 6 DSA. Further, according to ECJ case law, "knowledge" of illegal content in the meaning of Art. 6 DSA requires that the provider is able "to satisfy itself, without a detailed legal examination, that the communication of the content at issue is illegal and that removing that content is compatible with freedom of expression and information".[40] Art. 16 (3) DSA reinforces this case law.

Now imagine that a notice-giver has flagged specific content as illegal and the provider does not take action to remove the content because the provider cannot ascertain without a detailed legal examination that the content is in fact illegal. Imagine also that the decision remains unchanged after an internal review. If a court subsequently finds that the content was in fact illegal, should the court be able to award damages to the

[37] Cf. Holznagel, 'Zu starke Nutzerrechte in Art. 17 und 18 DSA', CR 2022, 594, 597 who correctly notes the discrepancy to § 3b (2) n. 1 of the German NetzDG. See also Bayer 'Procedural rights as safeguard for human rights in platform regulation', 17. January 2022, https://ssrn.com/abstract=4250224: 'DSA's procedural requirements appear to be more focused to diminish the power difference between online platforms and users, rather than to solve differences between users themselves'.

[38] Compensation claims were denied on the basis of current German law by OLG Frankfurt, Case 16 U 229/20, ZUM-RD 2022, 630.

[39] Art. 12 et seq. Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, OJ L 178/1.

[40] Case C-401/19, *Poland v. Parliament and Council*, para 91.

flagger because the provider failed to comply with the obligations under Art. 20 DSA? The answer is no. The provider is still entitled to rely upon the liability exemption of Art. 6 DSA.

If this is the standard of scrutiny for illegal content, the standard should not be any different if a violation of standard terms and conditions is at issue. As a consequence, Art. 20 (4) DSA only obliges the platform provider to take a "reasonable" review decision, not a "correct" one. As has been argued above, ISPs should be granted some authority in interpreting their own communication standards.[41] Recital 39 seems to reinforce this view, as it addresses judicial put back-decrees (only) in cases in which the provider erroneously considered content to be illegal.[42]

e) Measures Against Misuse

Whereas the BGH "Hate Speech" ruling pertained to sanctions that had been imposed by Facebook on a voluntary basis, the DSA will require online platforms to take action against misuses of their platform. Under Art. 23 (1) DSA, providers of online platforms shall suspend the provision of their services to recipients of the service that frequently provide manifestly illegal content.[43] The suspension obligation is temporary ("for a reasonable period of time") and subject to a prior warning of the user. When deciding upon the suspension, platforms must consider the absolute number of posted illegal content, the relative proportion of such content within a given time frame, the gravity of misuse, and possibly the intent behind it (if it can be identified).[44]

Art. 23 DSA does not limit a platform's ability to suspend a user or its service in other instances (i.e., for posting content that violates the platform's communication standards) or to permanently ban a user.[45] However, such measures must be set out in the provider's terms and conditions.[46] Also, since infringements of the communication standard will generally be less grave than frequent postings of manifestly illegal content, the principle of proportionality requires a prior warning for such suspensions and a

[41] Above at II. 4. c).

[42] 'This Regulation should therefore not prevent the relevant national judicial or administrative authorities from issuing, on the basis of the applicable Union or national law, an order to restore content, where such content was in compliance with the terms and conditions of the provider of the intermediary service but has been erroneously considered as illegal by that provider and has been removed'.

[43] According to recital 63 sent. 3, content should be considered manifestly illegal 'where it is evident to a layperson, without any substantive analysis, that the content is illegal'.

[44] Art. 23 (3) DSA.

[45] Cf. Art. 17 (1) (c), (d) DSA.

[46] Art. 14 (1) DSA, Art. 3 (1) Unfair Contract Terms Directive. Art. 23 (4) DSA only relates to illegal content.

proportionate response on the part of the platform provider.[47] Art. 23 (3) DSA may thus function as a guiding principle for voluntary suspension decisions taken by the platform.

f) Summary

In conclusion, the DSA and the BGH "Hate Speech" ruling follow similar approaches. Both allow providers to set their own communication standards and to restrict "lawful but awful" content. However, both also require providers to take into account the fundamental rights of users and third parties in shaping and applying their content moderation policies. Thus, the communication standard must be both transparent and proportionate.

As always, the devil lies in the details: Art. 14 DSA applies to all intermediary service providers, and Art. 17 and 20 DSA apply to host providers and online platforms respectively. In contrast, the BGH's "Hate Speech" ruling only applies to Facebook, and it is currently unclear whether and to what extent the decision may be transferable to other providers. Furthermore, the "Hate Speech" ruling is focused on the validity of standard terms and conditions, whereas Art. 14 (1) DSA merely requires providers to inform users about their policies in their terms and conditions.

An internal complaints mechanism must be available for a wide variety of sanctions by host providers under Art. 17 DSA, including for the demotion of content and restriction of monetization, irrespective of whether the sanction was imposed due to a violation of the law or of the provider's communication standards. In contrast, the BGH only had the opportunity to give guidance on the removal of content and suspension of user accounts in cases of a violation of the provider's communication standards. Further, while a host provider must notify an affected user of any measure taken both under the "Hate Speech" principles and under Art. 17 DSA, the requirements for that notice are set out in much more detail in Art. 17 (4) and (5) DSA. Perhaps more importantly, the BGH requires users to be heard before their account is suspended (even if temporarily or partially). In contrast, under Art. 17 (2) DSA, the notice is required at the latest from the date that the restriction was imposed. This difference is alleviated somewhat by the fact that users need to receive a warning before their account is suspended for frequent and manifest misuses of a platform – a rule that arguably also applies to minor infringements (*argumentum a maiore ad minus*).

Finally, an important distinction is that Art. 20 DSA makes the complaints mechanism accessible to notice-givers who have flagged illegal or harmful content. Unfortunately, the standard of internal review is unclear both under the DSA and under the

[47] See for current German law OLG Brandenburg, Case 4 U 1050/21 'Identitäre Bewegung', MMR 2022, 479 para 25; OLG Karlsruhe Case 10 U 17/20 para 144.

“Hate Speech” principles. As I have shown above, online platforms should not be obliged to undertake difficult legal analysis during the review process. Further, platforms should be granted some discretion in interpreting their own communication standards. Judicial review remains possible both regarding illegal content and in situations in which the communication standard is disproportionate or enforced arbitrarily.

IV. Impacts of the DSA on BGH Case Law

Now is the time to return to the question posed at the outset: What is the impact of the Digital Services Act on the case law of the German Federal Court of Justice? One option is to think of the “Hate Speech” ruling and the DSA as two separate strands running parallel to each other.[48] My suggestion, however, is that the BGH adapt its case law to the Digital Services Act. In its “Hate Speech” ruling, the BGH pulled a magic tric: Introducing freedom of speech safeguards via the judicial scrutiny of standard terms and conditions.[49] The route taken by the Court is understandable as the German Civil Code does not define the contractual obligations of online platforms. However, now that the legislature has spoken, statutory law should take priority over case law.

Furthermore, in the “Hate Speech” ruling, the BGH made use of the judicial control of standard terms and conditions to provide for a horizontal application of fundamental rights and require Facebook to implement procedural safeguards. Under § 307 (2) no. 1 BGB, standard terms are deemed to violate the principle of good faith if they are incompatible with the fundamental principles of the statutory provision from which they derogate. While the DSA does not primarily aim to regulate contractual relations, some of its provisions serve to specify the contractual relationship between providers and their users.[50] The provisions on the horizontal effect of fundamental rights, on transparency for content moderation policies, and on the corresponding procedural safeguards constitute fundamental statutory principles for the purposes of § 307 (2) n. 1 BGB, which any future review of standard terms would have to look to.[51]

[48] See Holznagel, ‘Zu starke Nutzerrechte in Art. 17 und 18 DSA’, [2022] CR 594, 598.

[49] Cf. the surprised reaction by Lutzi, ‘Plattformregulierung durch AGB-Kontrolle?’, Verfassungsblog, 30 July 2021, https://verfassungsblog.de/facebook-agb-kontrolle.

[50] Raue and Heesen, ‘Der Digital Services Act’, NJW 2022, 3537, para. 18. Even if the DSA provisions were to be considered non-contractual in nature, they would still have to be considered in the context of § 307 II n. 1 BGB, as non-contractual statutory law may provide fundamental principles for the purposes of § 307 II n. 1 BGB, cf. Wurmnest in MüKo BGB, 9th ed. 2022, § 307 BGB Rn. 76 ff.

[51] Lutzi, ‘Plattformregulierung durch AGB-Kontrolle?’, Verfassungsblog, 30. July 2021, https://verfassungsblog.de/facebook-agb-kontrolle.

What does this mean in practice? We need to distinguish between the "Hate Speech" principles regarding transparency on the one hand and procedural safeguards on the other.

Art. 14 (1) DSA only requires providers to inform their users about any restrictions of the service. The validity of those standard terms and conditions is thus a matter for the Unfair Contract Terms Directive and its national implementations, insofar as consumer contracts are concerned. Consequently, the BGH case law holding that service restrictions in standard terms and conditions must be transparent, lest the terms be invalid, still stands. With respect to business users, Art. 3 (1) (c), (3) P2B-Directive declares platform terms null and void unless they "set out the grounds for decisions to suspend or terminate or impose any other kind of restriction upon the provision of their online intermediation services to business users".

With respect to the procedural safeguards, the DSA requires changes to the BGH case law. First, the BGH should no longer require providers to hear users before the suspension of an account, since Art. 17 (2) DSA clearly follows a different route. The second change relates to the legal consequences of a failure to address procedural safeguards in the provider's terms and conditions. The BGH's line of argument in "Hate Speech" was as follows: Unless the statement of reasons and the complaints mechanism are stipulated in the ISP's terms and conditions, the provider is under no obligation to grant the complainant a right to review. This, however, would cause a significant imbalance in the parties' rights and obligations arising under the contract, thus rendering the provider's contract moderation clauses invalid. Under the DSA, the situation has changed: Host providers are required by law to provide their users with a statement of reasons for any sanction taken, and online platforms must implement an internal complaints mechanism. These obligations arise irrespective of whether the provider's terms and conditions refer to the procedural safeguards. Failure to name those safeguards in the terms and conditions thus does not cause a significant imbalance in the parties' rights and obligations. As a consequence, even if a provider fails to inform its users of such measures in its terms and conditions, content moderation decisions will remain in force. At most, a failure to inform users in accordance with Art. 14 (1) DSA could entail a claim to compensation under Art. 54 DSA.

It is open for debate whether the BGH could extend the "Hate Speech" principles on procedural safeguards to providers not covered by Art. 17 and 20 DSA. Certainly, a statement of reasons explaining why a particular sanction was taken would be helpful to users who are being banned by access providers. A complaints mechanism could also prove beneficial to users of sharehosting and webhosting services (who are not considered online platforms for purposes of Art. 20 DSA). According to recital 9 DSA, the regulation aims to fully harmonise the rules applicable to intermediary services, and Member States should not adopt or maintain additional national requirements relating

to the matters falling within the scope of this Regulation. On the other hand, the recital explains that this "should not preclude the possibility of applying other national legislation applicable to providers of intermediary services, in compliance with Union law". In my view, this still leaves an avenue for judicial scrutiny of unfair terms. However, the DSA clearly distinguishes between different types of providers and holds access, webhosting, and cloudhosting services to a lower standard. Remember, the fairness of contract terms is determined, amongst others, by their deviation from the statutory law (§ 307 (2) n. 1 BGB). It is therefore difficult to argue that the lack of a contractual stipulation of such safeguards renders the provider's terms and conditions invalid, even though the DSA does not require such safeguards.

V. Conclusion

Every new piece of secondary EU legislation raises new questions about its interfaces and compatibility with existing EU and Member States' law. The Digital Services Act is no exception. The BGH's "Hate Speech" ruling addresses similar topics as Art. 14, 17, and 20 DSA, albeit from the different angle of unfair contract terms control. As this article has shown, rather small modifications would allow the BGH to align its "Hate Speech" case law with the DSA requirements. This path is preferable to the application of two different sets of rules, one being the DSA requirements and the other being the BGH case law.

List of Abbreviations

ADBU	*Assam Don Bosco University*
AEUV	*Vertrag über die Arbeitsweise der Europäischen Union*
AfP	*Zeitschrift für das gesamte Medienrecht*
AGB	*Allgemeine Geschäftsbedingungen*
AI Act	*Artificial Intelligence*
BDSG	*Bundesdatenschutzgesetz*
BeckOK	*Beck'scher Online Kommentar*
BGH	*Bundesgerichtshof*
BGHZ	*Entscheidungen des Bundesgerichtshofes in Zivilsachen*
BVerfG	*Bundesverfassungsgericht*
BVerfGE	*Entscheidung des Bundesverfassungsgerichts*
Cf.	*Compare*
CFR	*Code of Federal Regulation*
CiTiP	*Centre of IT & IP Law*
CJEU	*Court of Justice of the European Union*
CR	*Computer und Recht*
CRi	*Computer Law Review International*
DAS	*Digital Services Act*
DMA	*Digital Markets Act*
DSGVO	*Datenschutz-Grundverordnung*
E.g.	*For example*
EC	*European Commission*
ECJ	*European Court of Justice*
ECLI	*Europäischer Rechtsprechungs-Identifikator*
ECtHR	*European Court of Human Rights*
EJRR	*European Journal of Risk Regulation*
Et at.	*And others*
Et seq.	*And the following*
EU	*European Union*
EuCML	*Journal of European Consumer and Market Law*
Eur. Rev. Priv. Law	*European Review of Private Law*
EUV	*Vertrag über die Europäische Union*
FAZ	*Frankfurter Allgemeine Zeitung*
FCC	*Federal Constitutional Court*
GDPR	*General Data Protection Regulation*
GPR	*Zeitschrift für das Privatrecht der Europäischen Union*

GRCh	*Grundrechtecharta*
GRUR	*Gewerblicher Rechtsschutz und Urheberrecht*
JIPITEC	*Journal of Intellectual Property, Information Technology and Electronic Commerce Law*
JZ	*JuristenZeitung*
KG	*Kammergericht*
MMR	*Zeitschrift für IT-Recht und Recht der Digitalisierung*
MStV	*Medienstaatsvertrag*
MüKo	*Münchener Kommentar*
NetzDG	*Netzwerkdurchsetzungsgesetz*
NJW	*Neue Juristische Wochenschrift*
NVwZ	*Neue Zeitschrift für Verwaltungsrecht*
OJ	*Online Journal*
OLG	*Oberlandesgericht*
OUP	*Oxford University Press*
OVG	*Oberverwaltungsgericht*
RW	*Rechtswissenschaften*
SZ	*Süddeutsche Zeitung*
U.S.	*United States*
UrhDaG	*Urheberrechts-Diensteanbieter-Gesetz*
UWG	*Gesetz gegen den unlauteren Wettbewerb*
ZEuP	*Zeitschrift für Europäisches Privatrecht*
ZfDR	*Zeitschrift für Digitalisierung und Recht*
ZfPW	*Zeitschrift für die gesamte Privatrechtswissenschaft*
ZRP	*Zeitschrift für Rechtspolitik*
ZUM	*Zeitschrift für Urheber- und Medienrecht*
ZUM-RD	*Zeitschrift für Urheber- und Medienrecht – Rechtsprechungsdienst*
ZusProt	*Additional Protocol*

Index of Authors

Florence G'Sell
Professor of Private Law at the University of Lorraine, Nancy and holder of the Digital, Governance and Sovereignty Chair at Sciences Po, Paris.

Ruth Janal, LL.M.
Professor of Private Law, Intellectual Property Law and Commercial Law at the University of Bayreuth.

Lea Katharina Kumkar
Assistant Professor of Civil Law, Commercial Law and Legal Issues of Digitalisation at Trier University and Member of the IRDT.

Martin Steinebach
Professor and Group Leadership of the area Media Security and IT Forensics at Frauenhofer SIT, Darmstadt.

Antje von Ungern-Sternberg
Professor of Public Law, Comparative Law, Law and Religion and Public International Law at Trier University as well as Director of the IRDT.

Mattias Wendel
Professor of Public Law, EU Law, International Law, Migration Law and Comparative Law at Leipzig University.

SCHRIFTEN DES IRDT

TRIER STUDIES ON DIGITAL LAW

Antje von Ungern-Sternberg (ed.)

Content Regulation in the European Union

The Digital Services Act

Illegal and (lawful, but) harmful content – most notably hate speech and fake news, but also violent videos, copyright infringement, or child pornography – is a crucial problem on digital platforms like Facebook, YouTube, TikTok and Twitter. The EU's 2022 Digital Services Act aims at tackling this problem by introducing an updated horizontal framework for all categories of content and activities on intermediary services. This raises several questions. How far do – national and European – free speech guarantees go? If hate speech can be banned to protect the victims' rights, how can the prohibition of fake news be justified? What is the remaining leeway of the platforms for private content moderation? Who is responsible for fighting and taking down illegal content? How can the victims of de-platforming, content takedowns or shadow banning claim their right to freedom of opinion? Finally, how will these legal responsibilities be enforced? These questions are addressed in the articles of the edited volume, proceeding from the 2022 Annual Conference of the Institute for Digital Law Trier (IRDT).