

Platform Practice and Non-Discrimination:
Conceptions of Equality, Systemic Risks, and Data Bias in the DSA
Felicitas Rachinger*

Contents

I. Introduction	124
II. Platforms and Non-Discrimination	125
III. Non-Discrimination in the DSA	127
IV. Individual Level: Non-Discrimination in Content Moderation.....	130
A. Obligations to Conduct Content Moderation in a Non-Discriminatory Manner	130
B. Societal Bias in Data	132
C. Technical Limitations	133
V. Systemic Level: Non-Discrimination and Risk Assessment	134
VI. Conclusion.....	136
VII. Bibliography.....	137

I. Introduction

Platforms have developed into important communicative facilities: We use them for connecting with others, sharing our ideas, and expressing our opinions. Rules and practices of platforms define what happens on a platform. These rules and practices are not free from discrimination. This is especially true as technical tools used for content moderation, such as hate speech detection tools or recommender systems, can transfer mechanisms of exclusion into digital spheres and might even reinforce

*Felicitas Rachinger, Department of Legal Theory and Future of Law, Universität Innsbruck, felicitas.rachinger@uibk.ac.at



them.¹ The EU legislator seems to recognize the issue and repeatedly emphasizes the relevance of "non-discrimination" in the Digital Services Act (DSA)². But what is "non-discrimination" when it comes to platforms? This paper addresses the conceptions of "non-discrimination" and equality within the DSA and asks how related provisions in the DSA can be put into practice.

The first section of the paper examines the connection between platforms and non-discrimination, discussing how public values are being incorporated into private orders (II.). The second section of the paper shows how the term "non-discrimination" and related terms can be understood under the DSA (III.). This is followed by an analysis of specific content moderation-related provisions on the individual (IV.) and the systemic level (V.)

II. Platforms and Non-Discrimination

As private entities, platforms can develop and implement their own rules and govern their platform to suit their interests, while still having to follow legal obligations.³ These ordering structures have become hybrid orders⁴, with private rules incorporating increasingly more public values. Public values, traditionally protected through fundamental rights, have gained significance for platforms, which reflects in

¹ Schwemer, Katzenbach, Dergacheva, Riis and Quintais, 'Impact of content moderation practices and technologies on access and diversity' (*D.6.3 Final Evaluation and Measuring Report*, 31 January 2023) <<https://doi.org/10.5281/zenodo.7705391>> accessed 10 August 2024; more general on possible harms of artificial intelligence and large language models: Bender, Gebru, McMillan-Major and Shmitchell, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (2021) FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610; Orwat, 'Diskriminierungsrisiken durch Verwendung von Algorithmen' (Berlin, 2019) 82ff; Schwartz, Vassilev, Greene, Perine, Burt and Hall, 'Towards a Standard for Identifying and Managing Bias in Artificial Intelligence' (2022) NIST Special Publication 1270.

² 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 (Digital Services Act), L 277/1.

³ Gorwa, 'What is platform governance?' (2019) *Information, Communication & Society* 854.

⁴ Kettemann, 'How Platforms Respond to Human Rights Conflicts Online: Best Practices in Weighing Rights and Obligations in Hybrid Online Orders' (Hamburg 2022).

academic discourses⁵ but also in court decisions⁶, new legislative acts such as the DSA. Even platforms themselves nowadays show an interest of implementing public values in their rules and practices⁷.

One of the key elements of “publicness” is the question of how non-discrimination is approached. Within the EU, Article 21 CFR⁸ sets the basis for the protection of the right to non-discrimination as a fundamental right. It prohibits discrimination based on a non-exhaustive⁹ enumeration of grounds such as “sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation”. While the provision primarily targets EU organs and member states when implementing Union law, several EU directives¹⁰ have been put in place to concretize Article 21 CFR and the scope of the right to non-discrimination in regard to private entities. In 2018, the CJEU confirmed the possibility of the horizontal application of Article 21 CFR in the case of *Egenberger*¹¹. The provision prohibits direct and indirect discrimination but does not constitute an absolute prohibition of discrimination, which instead can be justified following the principle of proportionality.¹²

Anti-discrimination obligations for private entities that are unrelated to discrimination at work are still limited within the EU. Two EU Directives aim at ensuring equal

⁵ Heldt, ‘Intensivere Drittwirkung: Die mittelbare Drittwirkung der Meinungsfreiheit in Öffentlichkeiten der digitalen Gesellschaft. Eine verfassungsrechtliche, rechtsvergleichende und interdisziplinäre Analyse’ (Tübingen 2023); Quintais, Appelman and Ó Fathaigh, ‘Using Terms and Conditions to Apply Fundamental Rights to Content Moderation’ (2023) German Law Journal 1.

⁶ See German BVerfG 22.05.2019, 1 BvQ 42/19; German BGH 29.07.2021, III ZR 179/20 and III ZR 192/20.

⁷ See for example: Meta, ‘Corporate Human Rights Policy’ (about.fb 2022) <<https://about.fb.com/wp-content/uploads/2022/10/Meta-Corporate-Human-Rights-Policy.pdf>> accessed 10 August 2024.

⁸ Charter of Fundamental Rights of the European Union (CFR).

⁹ Jarass, ‘Artikel 21’ in Jarass (ed.), *Charta der Grundrechte der Europäischen Union unter Einbeziehung der sonstigen Grundrechtsregelungen des Primärrechts und der EMRK: Kommentar*, 4th edn. (München 2021) para 9.

¹⁰ See Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services; Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

¹¹ CJEU C-414/16, *Egenberger*, ECLI:EU:C:2018:257.

¹² Jarass, ‘Artikel 21’ para. 26.

treatment in goods and services irrespective of gender¹³ and racial or ethnic origin¹⁴, but there is still no consensus to widen their scope to areas not yet covered.¹⁵ For platforms, however, recent EU legislation, namely the DSA, introduces further recognition of the right to non-discrimination as a public value that has to be considered in private orders of platforms. This relates to the fact that discriminatory platform practices are not unusual, and especially discrimination that has its basis in the use of algorithmic content moderation systems - such as hate speech detection tools or recommender systems - has shown to be especially subtle and hard to tackle¹⁶.

Treating content differently is part of the business model of platforms: recommender systems aim at keeping users on platforms, which contributes to the generation of income.¹⁷ Within this paper, discriminatory platform practices refer to platform practices that are usually based on bias and lead to the unequal treatment of persons and groups without reasonable justification.

III. Non-Discrimination in the DSA

The DSA includes several provisions on how the right of non-discrimination as public value is to be incorporated into the private orders of platforms. More broadly, the DSA puts a stronger focus on the protection of fundamental rights in digital spheres such as platforms.¹⁸ Article 1 DSA names the protection of fundamental rights as one of the main aims of the regulation. This is further concretized in Recital 3, which lists several particularly relevant fundamental rights, including the right to

¹³ See Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services.

¹⁴ Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

¹⁵ A proposal for a directive to extend the scope of EU-Anti-discrimination law has not found consensus within the EU: Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation, COM/2008/0426 final.

¹⁶ See FN 2; Oliva, Antonialli and Gomes, 'Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online' (2020), *Sexuality & Culture*, 700; Davidson, Bhattacharya and Weber, 'Racial Bias in Hate Speech and Abusive Language Detection Datasets', *Proceedings of the Third Workshop on Abusive Language Online* (2019) 25; Hoffmann, 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse' (2019) *Information, Communication & Society* 900.

¹⁷ Just, 'Kommunikationsplattformen: Entwicklung, Funktionen und Märkte' in Grabenwarter, Holoubek, Leitl-Staudinger (eds.), *REM 22: Regulierung von Kommunikationsplattformen* (Wien 2022) 1.

¹⁸ Quintais, Appelman and Ó Fathaigh (2023) *German Law Journal* 1.

non-discrimination, putting non-discrimination in the focus of platform regulation. As the right to non-discrimination is set out in Article 21 CFR and not initially meant to be applied to private entities¹⁹, a more detailed look at the provisions of the DSA is helpful to understand the scope and meaning of non-discrimination within the DSA.

Different terms – non-discrimination, equality, bias – are used within the DSA. While the terms are related, they cannot be used interchangeably. Bias is usually described as the pre-condition for discrimination, discrimination being the result of bias. When discrimination is a negative effect of bias, equality refers to equal treatment as a desirable goal.²⁰

First of all, an explicit reference to the right to non-discrimination (as set out in Article 21 CFR) can be found five times within the DSA²¹. References to “discriminatory” content or behavior can be found eight times²², seven of them in the recitals. Some of these references²³ mainly concern discrimination by other users or third parties, but not platforms themselves. As the regulation does not put obligations on these parties, it is only coherent that in these cases, the DSA does not refer to any legal provision such as Article 21 CFR. These references, therefore, do not give any indication of how the right to non-discrimination is to be applied to platforms. Other references²⁴ do relate to discriminatory platform practices but are generally quite vague.

“Equality” is a rather rarely used term within the DSA. It is included three times²⁵, once promoting the cooperation of the Board²⁶ with other institutions, including ones

¹⁹ See above II.

²⁰ The related term of “equity” is not used in the DSA. On different approaches and conceptions of equality and equity, see Minow, ‘Equality vs. Equity’ (2021) *American Journal of Law and Equality* 167.

²¹ Recital 3 DSA (aim of the regulation); Recital 52 DSA (notice and action mechanisms); Art 34 (1) and Recital 81 DSA (risk assessment); Art 48 (4) DSA (crisis protocols).

²² Recital 12 DSA (unlawful discriminatory content); Recital 26 DSA (good Samaritan clause); Recital 40 DSA (due diligence obligations); Recital 47 DSA (terms and conditions); Recitals 68 and 69 DSA (online advertising); Recital 58 and Article 20 (4) DSA (complaint-handling systems).

²³ Recital 12 DSA; Recital 40 DSA.

²⁴ Recital 26 DSA; Recital 47 DSA; Recitals 68 and 69 DSA; Recital 58 and Article 20 (4) DSA.

²⁵ Recital 95 DSA (Advertising systems); Recital 134 DSA (cooperation with other institutions); Art 47 DSA (codes of conduct for accessibility).

²⁶ Established under Section 3 / Article 61 DSA.

in the field of equality²⁷, and once referring to the equal participation of persons with disabilities, which shall be considered in any codes of conduct for accessibility that might be established under Art 47 DSA. The third reference relates to advertising systems used by very large online platforms (VLOPs) and very large online search engines (VLOSEs). The respective recitals urge these actors to conduct research into emerging risks, including real and foreseeable negative impacts on equality.²⁸ This refers back to the risk assessment obligations set out under Article 34 and 35 DSA, which in turn include an explicit reference to Article 21 CFR. Therefore, the use of the term “equality” in the context of advertising systems does not indicate a deviation from the understanding of non-discrimination that can be found in Article 21 CFR.

Lastly, the term “bias” can be found once in the DSA: Recital 94, concerning the risk mitigation of recommender systems. These measures should include measures to “prevent or minimize biases that lead to the discrimination of persons in vulnerable situations”. Bias as such is seen as a precondition for discrimination, and discrimination in turn as a result of bias. While the use of the term clarifies the relationship between bias and discrimination, it does not give any further legally relevant indication on how to apply the right to non-discrimination to platforms.

While different terms are used, they all relate to the concept of the right to non-discrimination that can be found in Article 21 CFR, therefore tackling direct as well as indirect discrimination.²⁹ Other than pointing to Article 21 CFR, references to discrimination are quite vague. This is especially problematic as no further indication is provided that relates to the process of balancing the right to non-discrimination with other fundamental rights also protected under the DSA, including fundamental rights of platforms. Finding ways to make use of Article 21 CFR in the context of platform regulation requires discussing DSA provisions in more detail.

As the above shows, terms related to non-discrimination are numerous within the DSA, but can be found mainly in recitals and not as often in the legal provisions themselves. While this shows that the EU legislator recognizes the potential of discrimination, it also shows there is still a lack of commitment to introducing legally binding non-discrimination obligations.

²⁷ Recital 134 DSA.

²⁸ Recital 95 DSA.

²⁹ See above II.

IV. Individual Level: Non-Discrimination in Content Moderation

A. Obligations to Conduct Content Moderation in a Non-Discriminatory Manner

A set of new obligations in the DSA concerns content moderation practices of platforms. This subsection aims at exploring how non-discrimination in content moderation is approached on an individual level within the regulation, focusing on how Article 21 CFR can be put into practice under the DSA.

One of the main new rules regarding content moderation relates to terms and conditions (T&Cs).³⁰ Within their T&Cs, providers of intermediary services have to be transparent about “any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the services”³¹. This rather lengthy wording can be summarized as a transparency requirement relating to the main content moderation rules and practices. It includes not only the policies that are the basis of their content moderation but also “any procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review”³². For providers of online platforms, additional transparency about recommender systems is required.³³ When applying and enforcing³⁴ these T&Cs, providers of intermediary services are required to consider the fundamental rights of all parties concerned.³⁵ This includes the right to non-discrimination, which is confirmed by the recitals, where it is stated that providers of intermediary services should act in a non-discriminatory manner when applying and enforcing the restrictions set out in T&Cs. Art 14 (4) DSA, therefore, requires the consideration of the right to non-discrimination in the application and enforcement of content moderation policies and practices that are set out in T&Cs. However, other fundamental rights have to be considered as well, including the fundamental

³⁰ Article 14 DSA. An in-depth analysis of the provision is provided by Quintais, Appelman and Ó Fathaigh (2023) German Law Journal 1.

³¹ Art 14 (1) DSA.

³² Art 14 (1) DSA.

³³ Art 27 DSA.

³⁴ Alongside the application and enforcement of the restrictions, the recitals also refer to the design thereof. Whether the legal obligations of Art 14 (4) extend to the design is controversial. For a restrictive interpretation see Maamar, ‘§ 4 Sorgfaltspflichten der Anbieter von Vermittlungsdiensten’ in Kraul (ed.), *Das neue Recht der digitalen Dienste (DSA)* (Baden-Baden, 2023) para. 47; for an extensive interpretation see Raue, ‘Artikel 14 DSA’ in Hofmann and Raue (eds.), *Digital Services Act* (Baden-Baden, 2023) paras. 74-76.

³⁵ Art 14 (4) DSA; Maamar, ‘§ 4 Sorgfaltspflichten der Anbieter von Vermittlungsdiensten’ para 46.

rights of providers of the intermediary service themselves. Art 14 (4) DSA particularly mentions freedom of expression, freedom and pluralism of the media, and also refers to “other fundamental rights and freedoms as enshrined in the Charter”. These fundamental rights might interfere with each other: For instance, the right to non-discrimination might interfere with the providers’ freedom to conduct a business.³⁶ Recital 153 DSA points to the application of the principle of proportionality to achieve a fair balance of the rights concerned. Considering the aim of the DSA to ensure a “safe, predictable and trusted online environment”³⁷ as well as the acknowledgment of the need of “responsible and diligent behavior by providers of intermediary services”³⁸ and also considering that the DSA recognizes digital environments to bear risks³⁹, it becomes apparent that special emphasis has to lie in the protection of the rights of users compared to those of platforms. Beyond these first indications, the DSA does not go into detail about this need for balancing.⁴⁰

Next to requirements relating to T&Cs in Article 14 DSA, Article 16 DSA requires the establishment of notice and action mechanisms. Such mechanisms shall allow users to notify providers of hosting services of illegal content. After receiving such notification, the providers are required to act in a “timely, diligent, non-arbitrary and objective manner”. While these terms might seem quite vague, it is again the recitals that provide background information on what this might include. The recitals refer to the protection of fundamental rights of all affected parties, including the right to non-discrimination.⁴¹

³⁶ For example, this might be the case if costly measures are required to apply and enforce T&Cs in a non-discriminatory manner.

³⁷ Art 1 (1) DSA.

³⁸ Recital 3 DSA.

³⁹ Recital 1 DSA.

⁴⁰ In addition to indications within the DSA itself, the balancing of fundamental rights has been subject to several CJEU cases, most notably the already mentioned case of Egenberger: CJEU C-414/16, Egenberger, ECLI:EU:C:2018:257. This case law can provide further information on the balancing of fundamental rights, but cannot be discussed in detail in this article. For further information, see for example Wischmeyer, ‘Grundrechtliche Bindung privater Plattformbetreiber unter dem EU Digital Services Act’ (2023); Quintais, Appelman and Ó Fathaigh (2023) German Law Journal 1; Frosio and Geiger, ‘Taking fundamental rights seriously in the Digital Services Act’s platform liability regime’ (2023) European Law Journal 31.

⁴¹ Recital 52 DSA.

To summarize the above, two of the main content moderation rules in the DSA oblige platforms⁴² to conduct content moderation in a non-discriminatory manner. However, several countervailing interests complicate putting this into practice. Two aspects seem especially problematic: the first is the question of whether platforms are responsible for discrimination based on societal bias in data. The second issue concerns technical limitations that might hinder platforms from mitigating or even detecting bias in their systems that may result in discrimination.

B. Societal Bias in Data

Algorithmic systems are often based on data from the past, meaning that within them societal bias and discriminating practices from the past are reflected.⁴³ One could ask: Why should platforms be better than society? Why should they be held accountable for discrimination that has its basis in societal bias in data sets? However, asking these questions misses that algorithmic systems do not only reflect societal biases but might also reinforce them.⁴⁴ Platforms play an active role in the development of such systems and decide how and when to use them. By doing so, they are not neutral actors but actively contributing to any possible algorithmic discrimination. As Anna-Katharina Mangold points out, the aim of anti-discrimination law is to change discriminatory societal practices.⁴⁵ Allowing the fact that bias or discriminatory practices are broadly present within society as a justification for discriminatory practices would undermine the aim of non-discrimination obligations. As a result, making use of Article 21 CFR in the context of societal bias in content moderation

⁴² Note: Article 16 DSA refers to providers of hosting services only, while Article 14 DSA refers to providers of intermediary services in general.

⁴³ Lopez, ‘Bias does not equal bias: a socio-technical typology of bias in data-based algorithmic systems’ (2021) Internet Policy Review 1.

⁴⁴ In the context of algorithmic content moderation systems: European Union Agency for Fundamental Rights (FRA), ‘Bias in Algorithms – Artificial Intelligence and Discrimination’ (Vienna, 2022) 49ff; Haimson, Delmonaco, Nie and Wegner, ‘Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas’ (2021) Proceedings of the ACM on Human-Computer Interaction 466/2021, 1; Prem and Krenn, ‘On Algorithmic Content Moderation’, in Werthner, Ghezzi, Kramer, Nida-Rümelin, Nuseibeh, Prem and Stanger (eds.), *Introduction to Digital Humanism* (Cham, 2024) 481 (489).

⁴⁵ Mangold, ‘Algorithmische Entscheidungssysteme: Eine neue Herausforderung für das (Antidiskriminierungs-)recht’ in Ulrich, Greif and Neuwirth (eds.) *Kritisches Rechtsdenken II: 10 Jahre Institut für Legal Gender Studies an der JKU: Linzer Schriften zu Gender und Recht 66* (Linz 2022) 48.

systems requires responsibility of platforms for discrimination resulting from such bias.

However, one cannot oversee that while platforms have a responsibility for discrimination stemming from societal bias in data, there might be several technical limitations to detecting or mitigating such bias.

C. Technical Limitations

Data bias is not easy to detect, and even if detected, not easy to mitigate.⁴⁶ It is therefore not unlikely that many algorithmic systems are in use that lead to discrimination, while the potential for discrimination is not known or no feasible technological solution is available. At the same time, platforms rely on algorithmic systems as an inherent part of their content moderation structures, not only to keep users interested in content but also to fulfill their legal obligations to handle user complaints and remove unlawful content within the required time periods.⁴⁷ As the DSA itself includes such obligations⁴⁸ and considers the use of algorithmic systems as a given⁴⁹, conducting content moderation in a non-discriminatory manner cannot mean a prohibition of the use of algorithmic systems that might lead to discrimination, especially in cases of subtle and undetected bias. While platforms cannot argue that they are not aware of the general potential of bias in algorithmic systems, forms of very subtle biases might still stay undetected. Article 14 and Article 16 DSA do not include obligations to implement processes to detect potential bias in their systems⁵⁰ and cannot provide sufficient protection from discrimination in cases of undetected bias in systems. To improve the protection of the right to non-discrimination, it is necessary to implement measures that go beyond the individual level and consider systemic implications. If algorithmic systems are heavily biased and influence communicative processes, the use of such systems has negative

⁴⁶ Several technical solutions for bias mitigation have been proposed and show the technical difficulties: Liu, Jia, Wei, Xu, Wang and Vosoughi, ‘Mitigating Political Bias in Language Models Through Reinforced Calibration’ (2021) The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21) 14857; Paola Lopez introduces a typology of bias that helps not only distinguishing different types of bias, but also suggests possible ways to address these biases: Lopez, Internet Policy Review 1 (21ff).

⁴⁷ Klonick, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’ (2018) Harvard Law Review 1598.

⁴⁸ Article 16 DSA.

⁴⁹ See for example Article 14 (1) DSA, Article 27 DSA.

⁵⁰ Even though the DSA does introduce transparency requirements related to automated content moderation systems in Art 14 DSA and Article 27 DSA, the potential of these provisions might be limited in cases of subtle bias.

implications on democratic processes and society at large. The DSA recognizes the systemic level with the introduction of systemic risk assessments but restricts its application on VLOPs and VLOSEs (see V.).

In cases of identifiable bias, there might be technical limitations to mitigating the potential of discrimination that lies within the system. In these cases, the application of DSA provisions depends on available, technical solutions as well as on the capabilities provider of intermediary or hosting services. As the right to non-discrimination has to be in balance with the protection of other fundamental rights, the freedom of providers to conduct a business has to be considered as well.⁵¹ As a result, while all platforms have a general obligation to conduct content moderation in a non-discriminatory manner, its concrete normative impact might vary significantly depending on the facts of an individual case.

V. Systemic Level: Non-Discrimination and Risk Assessment

With the introduction of a risk assessment procedure, the DSA allows for the protection of fundamental rights on a systemic level.⁵² Art 34 DSA considers “any actual or foreseeable negative effects for the exercise of fundamental rights, in particular the fundamental rights [...] to non-discrimination” as „systemic risk“. VLOPs and VLOSEs are required to regularly assess their services, in particular their content moderation systems, algorithmic systems and their T&Cs⁵³. As the provision explicitly refers to the right to non-discrimination, the DSA takes the important step of acknowledging discrimination as a potential problem that is inherent to content moderation systems. It allows for a systemic assessment that goes beyond individual cases of discrimination. Having in mind that discrimination in content moderation systems might lead to the silencing of marginalized communities and excluding opinions and perspectives from digital spheres, such discrimination negatively influences democratic processes and decision-making, requiring systemic approaches to platform non-discrimination regulation. The risk assessment requirement is therefore an important step in countering all forms of discrimination on VLOPs and VLOSEs and necessary to protect democratic processes.

⁵¹ Recital 52 DSA.

⁵² Mantelero, ‘Fundamental rights impact assessments in the DSA: Human rights and the risk-based approach of the new EU regulations on the digital society’ (Verfassungsblog, 1 November 2022) <<https://verfassungsblog.de/dsa-impact-assessment/>> accessed 10 August 2024; Maamar, ‘§ 4 Sorgfaltspflichten der Anbieter von Vermittlungsdiensten’ para. 213.

⁵³ Art 34 (2) DSA.

However, holistic systemic risk assessments require certain preconditions to allow for the detection and mitigation of risks. Regarding the right to non-discrimination, a reconceptualization of discrimination in digital constellations in light of intersectionality is required. The extensive enumeration of grounds of discrimination in Article 21 CFR requires a broad risk assessment, taking different forms of discrimination into account.⁵⁴ Without a profound understanding of discrimination and the different forms of discrimination that might appear on a systemic level, risk assessment cannot provide a successful tool to combat discrimination. This includes an understanding of intersectionality.⁵⁵ Currently, intersectionality is still rather rarely considered by the CJEU. Instead, the CJEU held in 2016 that there is “no new category of discrimination resulting from the combination of more than one of those grounds”⁵⁶. However, EU anti-discrimination law would allow for a broad interpretation of protected grounds of discrimination that considers intersectionality to a certain extent.⁵⁷ For holistic systemic risk assessment, it is essential to recognize this level and include intersectional discrimination in risk assessment processes.

In addition to an understanding of discrimination, it is necessary to have knowledge of the technical conditions that might lead to discrimination. This requires risk assessment teams being aware of how algorithmic bias can lead to algorithmic discrimination and how such bias can be detected as well as mitigated.⁵⁸

As a profound understanding of discrimination as well as technical aspects is required, interdisciplinary approaches are essential. An interdisciplinary team conducting the risk assessment should be supported through stakeholder consultations. While the DSA itself recognizes stakeholder consultations as a necessity in its recitals⁵⁹, no specific methodic requirements are in place. Therefore, while the DSA introduces an important first step with the acknowledgement of

⁵⁴ Kaesling, ‘Artikel 34 DSA’ in Hofmann and Raue (eds.), *Digital Services Act* (Baden-Baden, 2023) para 94.

⁵⁵ Allen, ‘An Intersectional Lens on Online Gender Based Violence and the Digital Services Act’ (Verfassungsblog, 01 November 2022) <<https://verfassungsblog.de/dsa-intersectional/>> accessed 10 August 2024.

⁵⁶ CJEU, C-443/15, ECLI:EU:C:2016:897, para. 80.

⁵⁷ Holzleithner, ‘Intersektionale (mehrdimensionale) Diskriminierung’, in Mangold and Payandeh (eds.), *Handbuch Antidiskriminierungsrecht* (Tübingen, 2022) 543 (570ff).

⁵⁸ See FN 39; for risk mitigation requirements see Art 35 DSA.

⁵⁹ Recital 90 DSA.

systemic levels of discrimination, the success of the provision lies within the persons conducting the risk assessment, therefore, VLOPs and VLOSEs themselves.

VI. Conclusion

The right to non-discrimination is a key right for a sustainable communication order in digital constellations. Platform practices and the use of algorithmic systems lead to discrimination and negatively influence democratic processes. Platforms need to be better in fostering a more inclusive, transparent, and fair environment for all users. The DSA recognizes the problem through focusing on Article 21 CFR as the main basis for any non-discrimination obligation within the DSA, but misses an opportunity to commit to strengthening the right to non-discrimination beyond vague commitments.

However, the DSA provides an important milestone in acknowledging the problem of algorithmic discrimination on a systemic level, not only the individual level. DSA non-discrimination obligations have to be implemented in content moderation for example via the inclusion in T&Cs (ex durante) and in risk assessment (ex ante) and allow for individual compensation⁶⁰ (ex post). By including these obligations, the DSA fills the right to non-discrimination with meaning in the digital era.

While the DSA introduces the obligation to conduct content moderation in a non-discriminatory manner, it offers limited guidance on balancing non-discrimination rights with other fundamental rights, including those of platforms. Practical limitations, especially the hard task of mitigating biases in datasets, prove to be an obstacle in fulfilling the full potential of non-discrimination obligations.

To approach the challenge effectively, platforms need to go beyond legal obligations. Implementing transparent content moderation criteria, standardizing regular audit processes of their content moderation systems, and introducing feedback loops and user engagement can contribute to strengthening marginalized communities. Additionally, conducting ongoing training of all team members working on content moderation issues is important: from human moderators to developers of AI systems to persons conducting the risk assessment.⁶¹ Such training would increase awareness

⁶⁰ Art 54 DSA.

⁶¹ For further recommendations on human rights-oriented content moderation, see Access Now, '26 recommendations on content governance: a guide for lawmakers, regulators, and company policy makers' (Access Now, 2020) <<https://www.accessnow.org/wp-content/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>> accessed 10 August 2024.

of the dangers of algorithmic discrimination and set the basis for actual non-discriminatory platform practices.

VII. Bibliography

Access Now, '26 recommendations on content governance: a guide for lawmakers, regulators, and company policy makers' (Access Now, 2020) <<https://www.accessnow.org/wp-content/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>> accessed 10 August 2024

Allen, Asha 'An Intersectional Lens on Online Gender Based Violence and the Digital Services Act' (Verfassungsblog, 01 November 2022) <<https://verfassungsblog.de/dsa-intersectional/>> accessed 10 August 2024

Bender, Emily M., Timnit, Gebru, McMillan-Major Angelina and Shmitchell, Shmargaret, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (2021) FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610

Davidson, Thomas, Bhattacharya, Debasmita and Weber, Ingmar, 'Racial Bias in Hate Speech and Abusive Language Detection Datasets', Proceedings of the Third Workshop on Abusive Language Online (2019) 25

European Union Agency for Fundamental Rights (FRA), 'Bias in Algorithms - Artificial Intelligence and Discrimination' (Vienna, 2022)

Frosio, Giancarlo and Geiger, Christophe, 'Taking fundamental rights seriously in the Digital Services Act's platform liability regime' (2023) European Law Journal 31

Gorwa, Robert, 'What is platform governance?' (2019) Information, Communication & Society 854

Haimson, Oliver L., Delmonaco, Daniel, Nie, Peipei, and Wegner, Andrea, 'Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas' (2021) Proceedings of the ACM on Human-Computer Interaction 466/2021, 1

Heldt, Amélie, 'Intensivere Drittwirkung: Die mittelbare Drittwirkung der Meinungsfreiheit in Öffentlichkeiten der digitalen Gesellschaft. Eine verfassungsrechtliche, rechtsvergleichende und interdisziplinäre Analyse' (Tübingen 2023)

- Hoffmann, Anna Lauren, ‘Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse’ (2019) *Information, Communication & Society* 900
- Holzleithner, Elisabeth, ‘Intersektionale (mehrdimensionale) Diskriminierung’, in Mangold, Anna-Katharina and Payandeh, Mehrdad (eds.), *Handbuch Antidiskriminierungsrecht* (Tübingen, 2022) 543
- Jarass, Hans D., ‘Artikel 21’ in Jarass, Hans D., (ed.), *Charta der Grundrechte der Europäischen Union unter Einbeziehung der sonstigen Grundrechtsregelungen des Primärrechts und der EMRK: Kommentar*, 4th edn. (München 2021)
- Just, Natascha, ‘Kommunikationsplattformen: Entwicklung, Funktionen und Märkte’ in Grabenwarter, Christoph, Holoubek, Michael and Leitl-Staudinger, Barbara (eds.), *REM 22: Regulierung von Kommunikationsplattformen* (Wien 2022) 1
- Kaesling, Katharina, ‘Artikel 34 DSA’ in Hofmann, Franz and Raue, Benjamin (eds.), *Digital Services Act* (Baden-Baden, 2023)
- Kettemann, Matthias C. (ed.), ‘How Platforms Respond to Human Rights Conflicts Online: Best Practices in Weighing Rights and Obligations in Hybrid Online Orders’ (Hamburg 2022)
- Klonick, Kate, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’ (2018) *Harvard Law Review* 1598
- Liu, Ruibo, Jia, Chenyan, Wei, Jason, Xu, Guangxuan, Wang, Lili and Vosoughi, Soroush, ‘Mitigating Political Bias in Language Models Through Reinforced Calibration’ (2021) *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)* 14857
- Lopez, Paola, ‘Bias does not equal bias: a socio-technical typology of bias in data-based algorithmic systems’ (2021) *Internet Policy Review* 1
- Maamar, Niklas ‘§ 4 Sorgfaltspflichten der Anbieter von Vermittlungsdiensten’ in Kraul, Torsten (ed.) *Das neue Recht der digitalen Dienste (DSA)* (Baden-Baden, 2023)
- Mangold, Anna-Katharina ‘Algorithmische Entscheidungssysteme: Eine neue Herausforderung für das (Antidiskriminierungs-)recht’ in Ulrich, Silvia, Greif, Elisabeth, and Neuwirth, Karin (eds.) *Kritisches Rechtsdenken II: 10 Jahre Institut für Legal Gender Studies an der JKU: Linzer Schriften zu Gender und Recht* 66 (Linz, 2022) 48
- Mantelero, Alessandro, ‘Fundamental rights impact assessments in the DSA: Human rights and the risk-based approach of the new EU regulations on the digital society’

(Verfassungsblog, 1 November 2022) <<https://verfassungsblog.de/dsa-impact-assessment/>> accessed 10 August 2024

Meta, ‘Corporate Human Rights Policy’ (about.fb 2022) <<https://about.fb.com/wp-content/uploads/2022/10/Meta-Corporate-Human-Rights-Policy.pdf>> accessed 10 August 2024

Minow, Martha, ‘Equality vs. Equity’ (2021) *American Journal of Law and Equality* 167

Oliva, Thiago Dias, Antonialli, Dennys Marcelo and Gomes, Alessandra, ‘Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online’ (2020), *Sexuality & Culture*, 700

Orwat, Carsten, ‘Diskriminierungsrisiken durch Verwendung von Algorithmen’ (Berlin, 2019)

Prem, Erich and Krenn, Brigitte, ‘On Algorithmic Content Moderation’, in Werthner, Hannes, Ghezzi, Carlo, Kramer, Jeff, Nida-Rümelin, Julian, Nuseibeh, Bashar, Prem, Erich, and Stanger, Allison (eds.), *Introduction to Digital Humanism* (Cham, 2024) 481

Quintais, João Pedro, Appelman, Naomi and Ó Fathaigh, Ronan ‘Using Terms and Conditions to Apply Fundamental Rights to Content Moderation’ (2023) *German Law Journal* 1

Raue, Benjamin, ‘Artikel 14 DSA’ in Hofmann, Franz and Raue, Benjamin (ed.), *Digital Services Act* (Baden-Baden, 2023)

Schwartz, Reva, Vassilev, Apostol, Greene, Kristen, Perine, Lori, Burt, Andrew and Hall, Patrick, ‘Towards a Standard for Identifying and Managing Bias in Artificial Intelligence’ (2022) NIST Special Publication 1270

Schwemer, Sebastian Felix, Katzenbach, Christian, Dergacheva, Daria, Riis, Thomas and Quintais, João Pedro, ‘Impact of content moderation practices and technologies on access and diversity’ (*D.6.3 Final Evaluation and Measuring Report*, 31 January 2023) <<https://doi.org/10.5281/zenodo.7705391>> accessed 10 August 2024

Wischmeyer, Thomas, ‘Grundrechtliche Bindung privater Plattformbetreiber unter dem EU Digital Services Act’ (2023)