

Dr. Victoria Ibold, München

## Die „Bekämpfung“ von „hate speech“ in einer wehrhaften Demokratie – Regulierung durch das NetzDG und den DSA

### Kommentar zum Vortrag von Prof. Dr. Armin Engländer<sup>1</sup>

Der folgende Kommentar zum Vortrag von Prof. Engländer ergänzt die Betrachtung der zur „Bekämpfung“ der Hasskriminalität vorgenommenen gesetzlichen Änderungen im materiellen Strafrecht: **A:** Um eine Eingrenzung des Begriffs der Hasskriminalität (I.), um die damit verbundenen Risiken für demokratische Strukturen im Kontext der sozialen Medien (II.) und um die Verortung weiterer Regulierungsvorhaben zum Schutz eben dieser Strukturen (III.). **B:** Um eine Darstellung des Vollzugsdefizits bei der Bekämpfung von Hasskriminalität (I.), der Verantwortungsstrukturen der Betreiber sozialer Medien nach dem NetzDG und dem Digital Services Act (DSA) (II.) sowie um die damit verbundenen Risiken durch ein „Overblocking“ (III.).

#### A. „Democracy Becomes Militant“

##### I. Begriff der Hasskriminalität

1937 beschrieb der ins US-amerikanische Exil geflüchtete jüdische Politologe und Verfassungsrechtler *Karl Loewenstein* die „politische Technik“,<sup>2</sup> mit welcher der Faschismus demokratische Strukturen in Deutschland und anderen europäischen Staaten zerstört hatte. Dazu zählte er u.a. ein Phänomen, welches sich dem heutigen Verständnis von hate speech<sup>3</sup> (als letztlich einem Synonym zur Hasskriminalität)<sup>4</sup> zuordnen lässt:

„General discontent is focussed on palpable objectives [...]. A technique of incessant repetition, of over-statements and over-simplifications, is evolved and applied. The different sections of the people are played off against one another.<sup>5</sup> [...] the vast armory of fascist technique includes the more subtle weapons of vilifying, defaming, slandering [...].<sup>6</sup>

„Die allgemeine Unzufriedenheit konzentriert sich auf greifbare Ziele [...]. Es wird eine Technik der ständigen Wiederholung, der Übertreibungen und der Vereinfachungen entwickelt und angewandt. Die verschiedenen Teile der Bevölkerung werden gegeneinander ausgespielt. [...]

<sup>1</sup> Der Text stellt eine erweiterte Fassung des Vortrags dar. Der Vortragsstil wurde weitgehend beibehalten.

<sup>2</sup> *Loewenstein*, APSR 31 (1937), 417, 422 und passim.

<sup>3</sup> Zur Begriffshistorie sowie weiteren Begriffsdefinitionen, die sich nur in der Formulierung, nicht aber in ihrer Grundaussage von den nachfolgenden Definitionen unterscheiden: *Sponholz*, in: Hoven (Hrsg.), *Das Phänomen „Digitaler Hass“*, Baden-Baden, 2023, S. 20 ff. (zitiert als: *Sponholz*, in: Hoven (Hrsg.), 2023, S.). Zudem *Krause*, *Hate Speech*, München, 2022, S. 1 ff. (zitiert als *Krause*, 2022, S.).

<sup>4</sup> Der Bundesgesetzgeber etwa verwendet anstatt des Begriffs der „hate speech“ oder der Hassrede den Begriff der Hasskriminalität, meint damit aber im Sinne von „hate speech“ eine „kommunikative“ Deliktsbegehung, vgl. BT-Dr. 19/17741, S. 1 (Entwurf eines Gesetzes zur Bekämpfung des Rechtsextremismus und der Hasskriminalität) sowie BT-Dr. 18/12356, S. 1 (Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerkdurchsetzungsgesetz – NetzDG]). Im Folgenden wird schwerpunktmäßig der Begriff der Hassrede verwendet werden, da er gegenüber dem Begriff der Hasskriminalität präziser gefasst ist.

<sup>5</sup> *Loewenstein*, APSR 31 (1937), 417, 423.

<sup>6</sup> *Loewenstein*, APSR 31 (1937), 638, 652.

zum umfangreichen Arsenal faschistischer Technik gehören auch die subtileren Waffen der Verunglimpfung, Diffamierung, Verleumdung.“

Legt man den vom Europarat 2022 in seiner Empfehlung an die Mitgliedstaaten zur Bekämpfung von hate speech verwandten Begriff eben dieser daneben, ergeben sich Parallelen:

„[...] hate speech is understood as all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as “race”, colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation.“<sup>7</sup>

„[...] unter Hassreden sind alle Arten von Äußerungen zu verstehen, die zu Gewalt, Hass oder Diskriminierung gegen eine Person oder eine Personengruppe aufstacheln, sie fördern, verbreiten oder rechtfertigen oder sie aufgrund ihrer tatsächlichen oder zugeschriebenen persönlichen Merkmale oder ihres Status verunglimpfen, wie z. B. "Rasse", Hautfarbe, Sprache, Religion, Nationalität, nationale oder ethnische Herkunft, Alter, Behinderung, Geschlecht, Geschlechtsidentität und sexuelle Ausrichtung.“

Nach beiden Ansätzen ist hate speech weniger mit einem Einzelkonflikt zwischen dem Sprechenden und dem Adressaten verbunden; es geht mehr um die Grundhaltung, dass es innerhalb einer Gesellschaft Gruppen mit geringerem Wert und geringeren Rechten gibt. In seiner Essenz beruht hate speech danach auch auf einer Art Kastendenken:<sup>8</sup> Hate speech äußert, wer der Überzeugung ist, dass ein Mitglied einer anderen Kaste auf Grund seiner Zugehörigkeit zu dieser weniger wert ist – weil es schwarz ist, weil es muslimischen oder jüdischen Glaubens ist, oder weil es eine Frau ist.<sup>9</sup> Treffend sind in diesem Zusammenhang daher auch die Begriffe der *Polarisierung* oder der *Spaltung*: Die Ablösung von einer gesellschaftlichen Zusammengehörigkeit (*Loewenstein* spricht von einer „common loyalty“<sup>10</sup>), die in einer Demokratie auch verbunden ist mit der Anerkennung der Gleichheit aller in ihrer Andersartigkeit; eine Entwicklung hin zu einer Zersplitterung in verschiedene gesellschaftliche Gruppen und einem „vilifying, defaming, slandering“<sup>11</sup> bestimmter Gruppen. Der Konflikt, welcher der hate speech zu Grunde liegt, ist somit nicht individueller, sondern systemischer Natur.

## II. Hate speech und soziale Medien

Hate speech und die damit verbundenen Risiken für ein demokratisches Gemeinwesen sind im Ergebnis nichts Neues. Neu hingegen sind die Mittel ihrer Verbreitung über das Internet, durch soziale Medien – über Facebook, Instagram, Tiktok, Twitter, Telegram etc. Die sozialen Medien mögen zwar nicht die Ursache von Hasskriminalität sein; Studien weisen inzwischen aber einen deutlichen kausalen Zusammenhang zwischen der Nutzung sozialer Medien in

<sup>7</sup> *European Council*, Recommendation CM/Rec(2022)16[1] of the Committee of Ministers to member States on combating hate speech, 20.5.2022, Appendix to Recommendation, Ziff. 1.1. Ähnlich bereits *European Council*, Recommendation No. R (97) 20 of the Committee of Ministers to Member States on "hate speech", 30.10.1997, Appendix to Recommendation No. R (97) 20, "Scope".

<sup>8</sup> Zu einer Analyse von Antisemitismus, Rassismus und Misogynie als Kastendenken *Wilkerson*, *Caste*, 2020, New York, S. 18, 29 f., 78 ff.

<sup>9</sup> *Sponholz*, in: Hoven (Hrsg.), 2023, S. 23 ff.; *Keiser*, ZRP 2010, 46; *Meddaugh/Jack*, *Journal of Mass Media Ethics* 24 (2008), 251, 254 ff.

<sup>10</sup> *Loewenstein*, *APSR* 31 (1937), 417, 432.

<sup>11</sup> „Verunglimpfung, Diffamierung, Verleumdung“, zum Nachweis siehe oben Fn. 6.

westlichen Demokratien und einer Zunahme von hate speech<sup>12</sup> sowie einer verstärkten Polarisierung im eben beschriebenen Sinne nach.<sup>13</sup> Diese Polarisierung findet statt in sog. „echo chambers“,<sup>14</sup> in welchen ähnliche Aussagen und Meinungen verstärkt und andersartige verdrängt werden.<sup>15</sup> Die Algorithmen der sozialen Medien können auch so gestaltet sein, dass die den Nutzern gezeigten Aussagen und Meinungen in „echo chambers“ immer „extremer“, immer polarisierter werden.<sup>16</sup>

Mit hate speech wird zudem ein silencing oder chilling effect verbunden;<sup>17</sup> Menschen würden darauf verzichten, sich auf eine bestimmte Art und Weise im Netz zu äußern oder sich überhaupt zu äußern. Die Annahme, dass Meinungsäußerungen aus Furcht vor den im Vergleich zur analogen Welt viel massiver geäußerten sowie breiter und langfristiger wahrnehmbaren Reaktionen bestimmter Gruppen im Netz unterlassen werden, ist plausibel.<sup>18</sup>

Dies alles mag dazu führen – und das ist die Befürchtung, die im Raum steht –, dass die polarisierenden Kräfte im öffentlichen Diskurs und in der öffentlichen Meinung an Dominanz gewinnen. Es geht in diesem Konflikt dann nicht mehr „nur“ um Rechte einzelner oder einzelner Gruppen, sondern auch darum, dass eine einseitige Dominanz der öffentlichen Meinung durch polarisierende Kräfte ein Risiko für die demokratische Verfasstheit der Gesellschaft als solches darstellt.<sup>19</sup> Dieses Risiko mag im zeitlichen Vergleich mit der Weimarer Republik heute geringer sein: Demokratische Strukturen sind gefestigter und die ökonomischen Verhältnisse besser. Dennoch ist weltweit wieder eine politische Blockbildung im Gange und bestimmte autoritär regierte Staaten stehen im Verdacht zu versuchen, mit den Mitteln auch der digitalen Hassrede demokratische Strukturen in der westlichen Welt zu unterwandern.<sup>20</sup>

<sup>12</sup> Vgl. die Übersichtsstudie von *Castaño-Pulgarín/Suárez-Betancur/Vega/López*, *Aggression and Violent Behavior* 58 (2021), 101608.

<sup>13</sup> *Schaub/Morisi*, *European Journal of Political Research* 59 (2020), 752: Zusammenhang zwischen Internetnutzung als Quelle für politische Informationen und der Wahl populistischer Parteien; *Allcott/Braghieri/Eichmeyer/Gentzkow*, *American Economic Review* 110 (2020), 629, 631: Deaktivierung von Facebook reduzierte die Polarisierung bzgl. politischer Themen in signifikanter Weise; *Bryson*, *International Journal of Sociology and Social Policy* 40 (2020), 99: Verfügbarkeit von politischer Information online erhöhte die Polarisierung signifikant; siehe zudem die Metastudie *Lorenz-Spreen/Oswald/Lewandowsky/Hertwig*, *Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide*, 14. April, 2022, S. 11 f. (zitiert als: *Spreen/Oswald/Lewandowsky/Hertwig*, 2022, S.) mit umfangreichen w.N.

<sup>14</sup> Dieser Begriff wurde erstmals verwendet von *Sunstein*, *Echo chambers*, Princeton, N.J, 2001.

<sup>15</sup> So v.a. bei Facebook und Twitter: *Cinelli/Francisci Morales/Galeazzi/Quattrociocchi/Starnini*, *Proc Natl Acad Sci USA* 118 (2021). Siehe zudem wiederum die Metastudie *Lorenz-Spreen/Oswald/Lewandowsky/Hertwig*, 2022, S. 12 f.

<sup>16</sup> So z.B. bei Facebook: *Kitchens/Johnson/Gray*, *MIS Quarterly* 44 (2020), 1619.

<sup>17</sup> Recommendation CM/Rec(2022)16[1] of the Committee of Ministers to member States on combating hate speech, 20.5.2022, 1; BT-Dr. 19/17741, S. 1; ähnl. *Ceffinato*, *JuS* 2020, 495, 497.

<sup>18</sup> Siehe die im Auftrag der Universität Leipzig erstellte Bevölkerungsumfrage; darin äußerten 42% der Befragten, aufgrund von Hassreden vorsichtiger bei eigenen Beiträgen im Internet zu sein, Universität Leipzig, *Hate Speech*, 2020, Hamburg. Zudem *Geschke/Klaßen/Quent/Richter*, in: Institut für Demokratie und Zivilgesellschaft, *#Hass im Netz*, 2019, S. 28 f. sowie die Interviewstudie von *Heuser/Witting*, in: Hoven (Hrsg.), *Das Phänomen „Digitaler Hass“*, Baden-Baden, 2023, S. 62.

<sup>19</sup> Recommendation CM/Rec(2022)16[1] of the Committee of Ministers to member States on combating hate speech, 20.5.2022, 1; *Ceffinato*, *ZStW* 132 (2020), 544, 549; *Lambrecht*, in: Hoven (Hrsg.), *Das Phänomen „Digitaler Hass“*, Baden-Baden, 2023, S. 10. Siehe zudem die Nachweise in Fn. 27, 28.

<sup>20</sup> Siehe z.B. die Recherchen der *Süddeutschen Zeitung* im Rahmen der „Vulkan Files“: *Cadenbach/Heubl/Kampf/Mascolo/Much/Muth/Sablowski/Weinmann/Wiegand/Ralf*, *Süddeutsche*

### III. Democracy Becomes Militant

„Democracy Becomes Militant [sic!]“.<sup>21</sup> Für *Loewenstein* lag die Antwort auf eine politische Technik, totalitäre Macht durch Ausnutzung demokratischer Strukturen zu erlangen, nicht in einer von Passivität geprägten „tolerant confidence of democratic ideology that in the long run truth is stronger than falsehood“.<sup>22</sup> Vielmehr entwickelte er – entgegen der noch zu Zeiten der Weimarer Republik herrschenden Auffassung eines demokratischen Relativismus<sup>23</sup> – die Idee einer „militant democracy“,<sup>24</sup> die der Ausnutzung demokratischer Strukturen etwas entgegen zu setzen vermag. Diese Idee wurde bei der Konzeption des Grundgesetzes übernommen<sup>25</sup> und prägt unter dem Begriff der wehrhaften, der streitbaren Demokratie die Rechtsprechung des Bundesverfassungsgerichts.<sup>26</sup>

Blickt man in die Begründungen der Gesetze, die in den letzten Jahren zur „Bekämpfung“ von hate speech auf den Weg gebracht wurden, wird diese in der Tat nicht nur unter Opferschutzgesichtspunkten, sondern auch als Risiko für die demokratische Verfasstheit gesehen. Zudem lesen sich die Gesetzesbegründungen als Ausdruck des Selbstverständnisses, dass die Bundesrepublik Deutschland als wehrhafte Demokratie verfasst ist. Hate speech berge eine „große Gefahr für das friedliche Zusammenleben einer freien, offenen und demokratischen Gesellschaft“,<sup>27</sup> sie berühre „wesentliche[n] Grundpfeiler der demokratischen pluralistischen Gesellschaft [...], die der Staat mit den ihm zur Verfügung stehenden Mitteln zu verteidigen hat.“<sup>28</sup> Eine „effektive Strafverfolgung“ sei dabei zentral.<sup>29</sup> Und so hat der nationale Gesetzgeber in den letzten Jahren die zahlreichen Straftatbestände

---

Zeitung 30.3.2023 Übersichtsartig mit weiteren Verweisen hauptsächlich auf journalistische Recherchen zudem *Wehner*, Osteuropa 67 (2017), 3, passim.

<sup>21</sup> *Loewenstein*, APSR 31 (1937), 417, 430.

<sup>22</sup> „tolerantem Vertrauen demokratischer Ideologie darauf, dass die Wahrheit auf Dauer stärker ist als die Lüge“, *Loewenstein*, APSR 31 (1937), 417, 424.

<sup>23</sup> *Radbruch*, in: Kaufmann (Hrsg.), Gustav Radbruch, Gesamtausgabe Bd. 12, Heidelberg, 1992, S. 97: „Dieser Relativismus, der die verschiedenen Parteauffassungen vom Allgemeinwohl nicht als beweisbar richtige oder unrichtige Feststellungen, sondern als unbeweisbare und unwiderlegbare Stellungnahmen betrachtet, ist die Grundlage demokratischen Denkens.“ Ebenso *Kelsen*, in: Jestaedt/Lepsius (Hrsg.), Verteidigung der Demokratie, Tübingen, 2006, S. 237: Bleibt sie [die Demokratie; Anm. der Verf.] sich selbst treu, muß sie auch eine auf Vernichtung der Demokratie gerichtete Bewegung dulden [...]. Siehe zu einem Überblick mit umfangreichen Nachweisen zudem *Papier/Durner*, AöR 128 (2003), 340, 343 ff.

<sup>24</sup> *Loewenstein*, APSR 31 (1937), 417; *Loewenstein*, APSR 31 (1937), 638 – jeweils passim. Eine ähnliche Idee entwickelte etwas später unter dem Begriff der „streitbaren Demokratie“ der ebenfalls ins Exil geflüchtete Soziologe *Karl Mannheim*, *Mannheim*, Diagnose unserer Zeit: Gedanken eines Soziologen, Zürich, 1951, S. 13 ff. (zitiert als: *Mannheim*, 1951, S.) (ursprünglich erschienen 1943 unter dem Titel „Diagnosis of our Time“, dort ist – wie bei *Loewenstein* – von der „militant democracy“ die Rede).

<sup>25</sup> Dazu gehört v.a. die Entscheidung für eine wertgebundene Demokratie; Art. 79 Abs. 3 GG entzieht die in Art. 1 und 20 GG niedergelegten verfassungsrechtlichen Grundstrukturen jeder politischen Disposition; zu weiteren verfassungsrechtlichen Elementen siehe die Übersicht bei *Papier/Durner*, AöR 128 (2003), 340, 347 ff.

<sup>26</sup> Der Schutz der „freiheitlichen demokratischen Grundordnung“ als Ausdruck des Gedankens der wehrhaften Demokratie erlaubt in den Grenzen des Grundgesetzes Einschränkungen einzelner Grundrechte; zum Begriff der freiheitlich demokratischen Grundordnung BVerfGE 2, 1 (11 ff.); BVerfGE 5, 85, 112, 140 ff.

<sup>27</sup> BT-Dr. 18/12356, 1.

<sup>28</sup> BT-Dr. 19/17741, S. 1.

<sup>29</sup> BT-Dr. 19/17741, S. 2; siehe daneben auch die Ausführungen von *Loewenstein*, wonach eine konsequente Anwendung des Strafrechts ein Baustein einer wehrhaften Demokratie sei, und zwar nicht nur bei gewaltsamem Widerstand oder Umsturzversuchen, sondern auch gegen „slander and libel“ – gegen Verleumdung und üble Nachrede; *Loewenstein*, APSR 31 (1937), 638, 645, 653.

des StGB, die dieses Phänomen zu erfassen vermögen, umgestaltet, um dieses aus seiner Sicht adäquater, d.h. umfassender und lückenloser, erfassen zu können.<sup>30</sup>

Indes: Der Regulierungsansatz für das Phänomen der hate speech – und damit ist nicht mehr nur der nationale, sondern inzwischen auch der unionale Gesetzgeber gemeint – geht weit über eine Steuerung mit den Mitteln des Strafrechts in einem engeren Sinne hinaus. Adressat strafbewehrter Verhaltensnormen ist nicht mehr nur der „Täter“ von hate speech; in die Verantwortung für rechtswidrige Inhalte genommen werden inzwischen auch die Betreiber sozialer Medien, die die Plattform für digitale Hasskriminalität bieten. Und für Zuwiderhandlungen seitens der Plattformbetreiber werden Bußgelder, also strafrechtliche Sanktionen in einem weiteren Sinne,<sup>31</sup> angekündigt. Im Folgenden sollen die Hintergründe dieses Regulierungsansatzes, die bußgeldbewehrten Pflichten der Betreiber sozialer Medien aus dem NetzDG und dem DSA sowie die damit verbundenen Folgen kritisch beleuchtet werden.

## B. NetzDG und DSA – Democracy becomes (too) militant?

### I. Vollzugsdefizit bei der Bekämpfung von Hasskriminalität

Die präventive Wirkung des Strafrechts hängt bekanntlich maßgeblich von der Wahrscheinlichkeit ab, dass eine Tat entdeckt, verfolgt und geahndet wird.<sup>32</sup> Im Bereich der hate speech ist diese Wahrscheinlichkeit bislang gering.<sup>33</sup> Das Internet ist zwar bei Weitem kein „anonymer“ Raum, der die Identität der Nutzer vollständig im Verborgenen lassen würde. Dennoch: Nutzer – insbesondere Täter von hate speech – treten oftmals nicht unter ihrem Klarnamen auf und die Daten,<sup>34</sup> mit Hilfe derer auf ihre Identität sowie ihr Handeln im Internet geschlossen werden könnte, sind nicht ohne Weiteres zugänglich<sup>35</sup> bzw. befinden sich bei den Betreibern sozialer Medien. Die Anzeigebereitschaft der Opfer ist äußerst zurückhaltend.<sup>36</sup>

### II. Verantwortung der Plattformbetreiber nach dem NetzDG und dem DSA

Angesichts dieses Vollzugsdefizits, konkret, um die Täteridentität mit Hilfe von Daten der Betreiber sozialer Medien aufdecken zu können, hat der Gesetzgeber in den letzten Jahren zunehmend die Plattformbetreiber in die Verantwortung genommen. Zunächst hat er in § 14 Abs. 2 Telemediengesetz eine Auskunftspflicht der Telemedienanbieter – und dazu gehören Betreiber sozialer Medien – gegenüber Strafverfolgungsbehörden statuiert. Für Zwecke der Strafverfolgung sind sie berechtigt und verpflichtet, Auskunft über Bestandsdaten, d.h. über personenbezogene Daten ihrer Nutzer, und unter erhöhten Voraussetzungen inzwischen auch über Nutzungsdaten zu geben.<sup>37</sup>

<sup>30</sup> Zu einer Übersicht der vorangehenden Beitrag von *Engländer*, [xx](#).

<sup>31</sup> Vgl. *Roxin/Greco*, Strafrecht Allgemeiner Teil I, München, 2020, § 2 Rn. 62, 98 f.

<sup>32</sup> Vgl. nur *Meier*, Strafrechtliche Sanktionen, Berlin, Heidelberg, 2019, 29 ff. m.w.N.

<sup>33</sup> So der Ausgangspunkt in BT-Dr. 19/17741, S. 1 f. Zu einer Analyse *Krause*, 2022, S. 5.

<sup>34</sup> Gesetzlich wird unterschieden zwischen Bestands- und Nutzungsdaten, vgl. §§ 22, 24, 2 Abs. 2 Nr. 2, 3 TTDSG.

<sup>35</sup> Siehe zu den Möglichkeiten, sich verdeckt im Internet zu bewegen (vpn-Clients, Thor-Browser) *Ceffinato*, ZStW 132 (2020), 544, 550.

<sup>36</sup> Eine FORSA-Umfrage für die Landesmedienanstalt NRW in den Jahren 2018, 2020 sowie 2022 hat ergeben, dass nur 1 % der Opfer von Hassrede bereits eine Strafanzeige erstattet haben, HATE SPEECH FORSA-STUDIE 2022, 2022, S. 5.

<sup>37</sup> Zur Rechtslage nach § 14 Abs. 2 TMG a.F.: *Nomos-BR/TMG-Müller-Broich*, § 14 Rn. 2, 5. Die Auskunftspflicht ist nunmehr in § 22 TTDSG (Telekommunikation-Telemedien-Datenschutz-Gesetz)

Diese auf Auskunftspflichten und damit auf eine weitgehend passive Rolle beschränkte Verantwortung der Betreiber sozialer Medien im Umgang mit Hassrede hat sich mit Einführung des Netzwerkdurchsetzungsgesetzes – dem NetzDG – 2017 geändert:<sup>38</sup> Die Betreiber wurden verpflichtet: ein umfassendes Beschwerdemanagement einzurichten,<sup>39</sup> offensichtlich rechtswidrige Inhalte i.S.d. NetzDG<sup>40</sup> innerhalb von 24h und sonst rechtswidrige Inhalte innerhalb von sieben Tagen nach Eingang der Beschwerde zu löschen<sup>41</sup> sowie zuletzt die Beschwerdeführer sowie von der Löschung betroffene Nutzer entsprechend zu informieren und Entscheidungen zu begründen.<sup>42</sup> Zudem wurden sie verpflichtet und ermächtigt, gegenüber Nutzern Auskunft zur Durchsetzung zivilrechtlicher Ansprüche zu erteilen.<sup>43</sup> Flankiert werden diese Pflichten mit der Ankündigung von Bußgeldern für den Fall des Zuwiderhandelns,<sup>44</sup> die nach Maßgabe der §§ 30, 130 OWiG auch gegenüber den Unternehmen selbst verhängt werden können.

Mit Wirkung zum Februar 2022 sind Änderungen des NetzDG in Kraft getreten,<sup>45</sup> die am Vollzugsdefizit ansetzen. Eingeführt wurde eine Meldepflicht: Die Betreiber sozialer Medien haben bestimmte von Nutzern angezeigte rechtswidrige Inhalte,<sup>46</sup> denen eine besonders gefährliche Wirkung auf das demokratische System und die öffentliche Ordnung zugeschrieben wird,<sup>47</sup> an das Bundeskriminalamt zu melden; von dort aus soll dann die Strafverfolgung durch die zuständigen Stellen veranlasst werden können. Diese Änderung des NetzDG ist jedoch derzeit ausgesetzt, da die Meldepflicht nach Ansicht des VG Köln gegen Unionsrecht verstößt.<sup>48</sup>

Das NetzDG in seiner ursprünglichen Fassung setzt, abgesehen von der Auskunftspflicht gegenüber den Nutzern, nicht am beschriebenen Vollzugsdefizit an und verpflichtet die Anbieter sozialer Medien auch nicht, Maßnahmen zu ergreifen, damit rechtswidrige Inhalte i.S.d. NetzDG gar nicht erst veröffentlicht werden können. Es geht also nicht um Verhaltensnormen mit einer unmittelbar präventiven Zwecksetzung, die eine Rechtsgutsverletzung verhindern, sondern um solche, die eine Tatbeendigung bewirken und damit die Intensität der Rechtsgutsverletzung vermindern sollen.<sup>49</sup>

---

iVm § 100j StPO geregelt. Eine Herausgabe von Nutzungsdaten ist danach unter den erhöhten Voraussetzungen des § 24 TTDSG möglich.

<sup>38</sup> Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG) vom 1.9.2017, in Kraft getreten am 1.10.2017 (BGBl. I S. 3352).

<sup>39</sup> § 3 Abs. 1 NetzDG.

<sup>40</sup> Rechtswidrige Inhalte i.S.d. NetzDG sind gem. § 1 Abs. 3 NetzDG solche, die den Tatbestand der §§ 86, 86a, 89a, 91, 100a, 111, 126, 129 bis 129b, 130, 131, 140, 166, 184b, 185 bis 187, 189, 201a, 241 oder 269 StGB erfüllen und nicht gerechtfertigt sind.

<sup>41</sup> § 3 Abs. 2 Nr. 2, 3 NetzDG.

<sup>42</sup> § 3 Abs. 2 Nr. 5 NetzDG.

<sup>43</sup> § 14 Abs. 3 TMG i.d.F. vom 1.10.2017.

<sup>44</sup> § 4 NetzDG.

<sup>45</sup> Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität, BGBl. I, S. 441, die Änderungen des NetzDG sind zum 1.2.2022 in Kraft getreten, vgl. Art. 10 Abs. 2.

<sup>46</sup> Gem. § 3a Abs. 2 Nr. 3 NetzDG sind dies Inhalte, welche die §§ 86, 86a, 89a, 91, 126, 129 bis 129b, 130, 131 oder 140, 184b StGB erfüllen und nicht gerechtfertigt sind. Erfasst sind auch Inhalte, die den § 241 StGB erfüllen, sofern sich die Bedrohung auf ein Verbrechen gegen das Leben, die sexuelle Selbstbestimmung, die körperliche Unversehrtheit oder die persönliche Freiheit bezieht.

<sup>47</sup> BT-DrS 19/17741, S. 17.

<sup>48</sup> LG Köln BeckRS 2022, 3081.

<sup>49</sup> Zu einer ähnlichen Analyse *Ceffinato*, ZStW 132 (2020), 544, 560 f.

Im November 2022 ist nun mit dem Digital Services Act (DSA)<sup>50</sup> eine unionsweite Regelung in Kraft getreten, die im Hinblick auf rechtswidrige Inhalte<sup>51</sup> auf ein ähnliches Regelungskonzept wie das NetzDG setzt. Ab Februar 2024 sollen für Anbieter von „Online-Plattformen“ u.a. folgende Pflichten gelten: Verpflichtung zur Einrichtung eines Melde- und Abhilfeverfahrens zur Meldung rechtswidriger Inhalte,<sup>52</sup> zu einer Entscheidung über die gemeldeten Informationen „zeitnah, sorgfältig, frei von Willkür und objektiv“<sup>53</sup>, versehen mit einer entsprechenden Begründung gegenüber den betroffenen Nutzern<sup>54</sup> und zuletzt zu einer Meldung des Verdachts auf begangene oder zukünftige Straftaten in Bezug auf das Rechtsgut Leben oder „Sicherheit“ einer Person.<sup>55</sup> Sehr große Online-Plattformen<sup>56</sup> werden zuletzt verpflichtet, eine umfangreiche Risikobewertung und Risikominderungsmaßnahmen zu ergreifen;<sup>57</sup> der DSA stellt jedoch klar, dass sich daraus keine allgemeine Verpflichtung zur proaktiven Kontrolle von Inhalten ergibt.<sup>58</sup>

Diese „reaktive“ Verantwortung der Netzbetreiber ist nicht unumstritten. Es geht dabei weniger um die damit verbundene Beeinträchtigung ihrer (auch grundrechtlich geschützten) ökonomischen Interessen<sup>59</sup> – soziale Medien sind die Plattform zur Verbreitung von hate speech und deren Zunahme lässt sich auch auf die zunehmende Verbreitung digitaler Medien zurückführen.<sup>60</sup> Wenn nun also von der Tätigkeit der Betreiber sozialer Medien Risiken ausgehen, erscheint es gerechtfertigt, sie im beschriebenen Sinne in die Verantwortung zu nehmen, selbst wenn sie nicht die „Verursacher“ digitalen Hasses sind. Die Kritik bezieht sich vielmehr darauf, dass die vom NetzDG statuierte reaktive Verantwortung faktisch zu einem „Overblocking“ führe, dass also nicht nur rechtswidrige, d.h. strafrechtsrelevante Inhalte gelöscht werden, sondern auch solche, die diese Grenze noch nicht überschreiten, sondern vielmehr vom Grundrecht auf freie Meinungsäußerung gedeckt sind.<sup>61</sup>

### III. „Overblocking“ – Democracy becomes (too) militant?

Eine Teilevaluation zu den Auswirkungen des NetzDG aus dem Jahre 2021 ist zum Ergebnis gelangt, dass die kurzen Lösungsfristen und die bei Zuwiderhandlung angekündigten hohen Bußgelder des NetzDG wohl eine Ausweichbewegung der Betreiber sozialer Medien in Richtung eines Overblocking herbeigeführt haben. Diese Ausweichbewegungen bestünden in der proaktiven Löschung von Inhalten v.a. mit Hilfe automatisierter Filter sowie in einer

<sup>50</sup> Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19.10.2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste) in Kraft getreten zum 16.11.2022 mit Geltung ab dem 17.2.2024.

<sup>51</sup> Der *Digital Services Act* gilt allgemein für digitale Vermittlungsdienste und soll ein sicheres, vorhersehbares und vertrauenswürdiges Online-Umfeld schaffen, vgl. Art. 1 Abs. 1 DSA sowie die Erwägungsgründe 3-5; es geht nicht nur um Vorgaben zum Umgang mit rechtswidrigen Inhalten, der DSA ist somit breiter angelegt als das NetzDG.

<sup>52</sup> Art. 16 Abs. 1 DSA; die Rechtswidrigkeit eines Inhalts bestimmt die Verordnung nicht selbst, rechtswidrig sollen Inhalte sein, deren Informationen „nicht im Einklang mit dem Unionsrecht oder dem Recht eines Mitgliedstaats stehen“, Art. 1 lit. h DSA.

<sup>53</sup> Art. 16 Abs. 6 DSA.

<sup>54</sup> Art. 17 DSA.

<sup>55</sup> Art. 18 DSA.

<sup>56</sup> Zum Begriff der Online-Plattform bzw. der großen Online-Plattform Art. 1 lit. i, Art. 33 DSA.

<sup>57</sup> Art. 34, 35 DSA.

<sup>58</sup> Art. 8 DSA – „Keine allgemeine Verpflichtung zur Überwachung oder aktiven Nachforschung“.

<sup>59</sup> Vgl. dazu *Frenzel*, JuS 2017, 414, 415 f.

<sup>60</sup> Siehe oben A.II.

<sup>61</sup> *Papier*, NJW 2018, 3030; *Guggenberger*, ZRP 2017, 98, 100; *Kalscheuer/Hornung*, NVwZ 2017, 1721, 1723; *Koreng*, GRUR-Prax 2017, 203, 204; *Nolte*, ZUM 2017, 552, 555 ff. A.A. *Eifert/Gostomzyk*, Netzwerkrecht, 2018, S. 9 f., 35 f.

reaktiven Löschung von Inhalten primär auf Grundlage Allgemeiner Geschäftsbedingungen (AGB) der Sozialen Medien („Flucht in die AGB“).<sup>62</sup> Legt man diese Annahme zu Grunde, führen nicht Löschungen auf Grund gesetzlicher Vorgaben nach dem NetzDG zu einem Overblocking, wohl aber die von den Betreibern sozialer Medien ergriffenen Maßnahmen einer Selbstregulierung. Für den in der Grundstruktur mit dem NetzDG vergleichbaren DSA ist zu erwarten, dass sich auch durch diesen Regulierungsrahmen das Problem eines mittelbar verursachten Overblockings stellt.

### 1. Proaktive Löschungen durch automatisierte Filter

Die oben erwähnte Studie hat gezeigt, dass proaktive Löschungen durch automatisierte Filter bei den großen sozialen Medien einen überragenden Anteil bei der Löschung von Inhalten einnehmen, weit vor reaktiven Löschungen auf Grund ihrer AGB und erst Recht vor reaktiven Löschungen auf Grund des NetzDG.<sup>63</sup>

Automatisierte Filter ersetzen den menschlichen Einsatz beim Herausfiltern bzw. beim Bewerten rechtswidriger Inhalte, was angesichts der großen Masse an digitalen Kommunikationsinhalten wohl die „kostengünstigere“ Variante ist. So wie Betreiber sozialer Medien gerade aus dem US-amerikanischen Bereich schon seit Jahren recht erfolgreich dabei sind, Nacktbilder in ihren Netzwerken herauszufiltern und zu sperren, setzen sie nun also Programme der Künstlichen Intelligenz<sup>64</sup> auch bei der Suche nach rechtswidrigen Kommunikationsinhalten ein. Aber: Sprache ist ungleich komplexer als Nacktheit, entsprechend schwieriger zu filtern. Und so geschieht es, dass automatische Filterfunktionen – bewusst oder unbewusst – zu sensibel ausgestaltet sind. Es wurde bspw. bekannt, dass die Videoplattform Tiktok Kommentare für andere Nutzer (nicht aber für den Kommentierenden)<sup>65</sup> blockiert, die u.a. folgende Begriffe enthalten: „Auschwitz, gay, Heterosexuelle, homosexuell, LGBTQ, LGBTQI, Nationalsozialismus, Peng Shuai, Porno, Pornografie, Prostitution, queer, schwul, Sex, Sexarbeit, Sklaven, Terroristen“.<sup>66</sup> Dass in einem solchen Fall auch Inhalte blockiert werden, die selbstverständlich von der freien Meinungsäußerung gedeckt sind, ist offensichtlich.

### 2. Reaktive Löschungen – Flucht in die AGB

Betrachtet man den – geringen – Anteil reaktiver Löschungen auf Grund von Nutzerbeschwerden, erfolgt wiederum ein weit überwiegender Anteil nicht auf Grundlage des vom NetzDG vorgegebenen strafrechtlichen Maßstabes, sondern auf Grundlage der AGB der

<sup>62</sup> *Liesching/Funke*, Das NetzDG in der praktischen Anwendung, 2021, Berlin, S. 359 ff. sowie passim (zitiert als *Liesching/Funke*, 2021, S.).

<sup>63</sup> Bspw. seien bei YouTube im Jahr 2020 94,36% der Löschungen aufgrund automatischer Erkennung erfolgt, bei Facebook im 4. Quartal 2020 97,1 %, *Liesching/Funke*, 2021, S. 360.

<sup>64</sup> Zur Technik der Künstlichen Intelligenz *Ibold*, ZStW 134 (2022), 504, 509 ff.

<sup>65</sup> Sog. Shadowbanning, vgl. *Meineck*, TikTok hat in Deutschland heimlich Kommentare blockiert, 2022, <https://netzpolitik.org/2022/geheime-wortfilter-tiktok-hat-in-deutschland-heimlich-kommentare-blockiert/>.

<sup>66</sup> Dies ergaben Recherchen von Recherchen von NDR, WDR und tagesschau; dazu: *Eckert/Felke/Grill/Milling/Vitlif*, TikTok nutzt in Deutschland Wortfilter, 2022, <https://www.tagesschau.de/investigativ/tik-tok-begriffe-blockade-101.html>.



Betreiber.<sup>67</sup> Diese sind überwiegend weiter ausgestaltet, erlauben also Löschungen von Inhalten, die nicht strafbar sind.<sup>68</sup>

Die These, dass Löschungen auf Grund eigener AGB Ausweichbewegungen im Hinblick auf die engen Vorgaben des NetzDG sind, erscheint plausibel. Dafür spricht, dass die in der erwähnten Studie in den Blick genommenen Betreiber bei Beschwerden stets zunächst nach ihren eigenen Maßstäben und erst dann nach dem NetzDG prüfen,<sup>69</sup> dass die weit überwiegende Anzahl der Löschungen innerhalb der engen Frist von 24 Stunden für „offensichtlich rechtswidrige Inhalte“ stattfindet und nicht innerhalb der für Zweifelsfälle vorgesehenen längeren Frist von sieben Tagen;<sup>70</sup> und zuletzt ist zu beachten, dass eine „Flucht in die AGB“ auch kein rechtswidriges Verhalten ist; der BGH sieht die Anbieter eines sozialen Netzwerks als grundsätzlich berechtigt an, den Nutzern in AGB die Einhaltung „objektiver, überprüfbarer Kommunikationsstandards vorzugeben, die über die gesetzlichen Vorgaben hinausgehen.“<sup>71</sup>

### C. Ausblick

*Loewenstein* beschrieb 1937 auch das grundsätzliche Dilemma, das mit einem Streben nach einer wehrhaften Demokratie verbunden ist. Welches sind die *legitimen* Mittel, die ein demokratischer Staat totalitären Bewegungen entgegensetzen kann, ohne selbst totalitäre Züge aufzuweisen?<sup>72</sup> Konkret in Bezug auf Hasskriminalität: „[...] the border-line between unlawful slander and justified criticism as lawful exercise of political rights is exceedingly dim.“<sup>73</sup> Eine Demokratie unterscheidet sich von totalitären Strukturen schließlich dadurch, dass sie gesellschaftlichen Zusammenhalt gerade nicht mit Rechtszwang durchsetzt.<sup>74</sup>

Wann wird also eine Demokratie *too* militant – zu wehrhaft, indem sie unliebsame Meinungen unverhältnismäßig einschränkt? Wie ist damit umzugehen, dass das NetzDG augenscheinlich mittelbar dazu führt, dass die Meinungsfreiheit unverhältnismäßig beeinträchtigt wird?

Auch im digitalen Raum muss Rechtsvollzug möglich sein. Regulierung darf Rechtskontrolle sein, sie soll dazu beitragen, dass rechtswidrige Äußerungen gelöscht und sanktioniert werden. Und dafür werden die Betreiber Sozialer Medien berechtigterweise in die Verantwortung genommen, weil sie die Plattform für Hassrede bieten. Gerade die massenhaft vorgenommenen proaktiven Löschungen auf Grund automatisierter Filter überschreiten aber die Grenze zur Gesinnungskontrolle, indem lediglich unliebsame oder unbequeme Ansichten und Meinungen unterdrückt werden. Auch wenn dies mittelbar auf Vorgaben des Gesetzgebers zurückgeht, sollte aber an der Verantwortung der Betreiber Sozialer Medien

<sup>67</sup> Nachweis *Liesching/Funke*, 2021, S. 127 ff.

<sup>68</sup> *Liesching/Funke*, 2021, S. 123 ff.; *Schwartzmann/Mühlenbeck*, ZRP 2020, 170.

<sup>69</sup> *Liesching/Funke*, 2021, S. 121, 126 ff.

<sup>70</sup> *Liesching/Funke*, 2021, S. 131 ff.

<sup>71</sup> BGH NJW 2021, 3179, 3188; BGH ZUM-RD 2021, 612, 625.

<sup>72</sup> „Democracy stands for fundamental rights, for fair play for all opinions, for free speech, assembly, press. How could it address itself to curtailing these without destroying the very basis of its existence and justification?“ (Die Demokratie steht für Grundrechte, für Fairness gegenüber allen Meinungen, für Rede-, Versammlungs- und Pressefreiheit. Wie sollte sie diese einschränken, ohne die Grundlage ihres Bestehens und ihrer Rechtfertigung zu zerstören?) *Loewenstein*, APSR 31 (1937), 417, 430 f.

<sup>73</sup> „[...] die Grenze zwischen rechtswidriger Verleumdung und berechtigter Kritik als rechtmäßiger Ausübung politischer Rechte ist äußerst dünn“, *Loewenstein*, APSR 31 (1937), 638, 653.

<sup>74</sup> Vgl. zu diesem Gedanken *Mannheim*, 1951, S. 17 f.; *Böckenförde*, Recht, Staat, Freiheit, 1991, Frankfurt am Main, S. 112 f.

festgehalten werden und diese dürfen ihre Verantwortung nicht auf die Technik abwälzen. Der DSA bietet nun eine neue Möglichkeit, demokratische Strukturen wehrhaft zu machen, ohne die Grenze zur Gesinnungskontrolle zu überschreiten.