KI-Lösungen gegen digitale Desinformation: Rechtspflichten und -befugnisse der Anbieter von Social Networks

Lena Isabell Löber

Zusammenfassung

Digitale Desinformation und algorithmenbasierte Manipulationen wie Social Bots und Deepfakes bergen Risiken für zahlreiche individuelle und gesellschaftsbezogene Rechtsgüter und fordern insbesondere den individuellen und öffentlichen Meinungsbildungsprozess heraus. KI-Lösungen sind essenzielle Instrumente, um solche schädlichen Inhalte und Manipulationstechniken in Social Networks zu detektieren. Die mit ihrem Einsatz verbundenen Risiken für Kommunikationsgrundrechte und Meinungspluralität sind durch manuelle Nachkontrollen automatisiert ermittelter Treffer und einen verfahrensbasierten Grundrechtsschutz einzuhegen. Zudem sind schärfere Transparenzvorgaben und Aufsichtsstrukturen erforderlich, um den Risiken der technisch-organisatorischen Gestaltungs- und Entscheidungsmacht großer Anbieter von Social Networks z. B. im Rahmen der algorithmischen Empfehlungssysteme zu begegnen. Im Hinblick auf den erheblichen Wissensvorsprung der Anbieter hat das nationale Gesetzesrecht mit neuen Regelungen im Medienstaatsvertrag und Netzwerkdurchsetzungsgesetz zu einigen Transparenzsteigerungen geführt, die jedoch gerade beim Themenkomplex Desinformation weitestgehend vage bleiben. Demgegenüber sind auf EU-Ebene im Rahmen der Entwürfe für die KI-Verordnung und den Digital Services Act neben dem deutschen Recht zum Teil sehr ähnlichen Regelungen auch weitergehende Pflichten vorgesehen, die einen wichtigen Beitrag zu einem ganzheitlicheren Ansatz im Umgang mit digitaler Desinformation leisten könnten.

1 KI-Lösungen als unverzichtbare Instrumente zur Eindämmung von Desinformation

Die algorithmenbasierte Informationssteuerung global agierender Social Networks nimmt Einfluss auf die Verbreitung und Wahrnehmbarkeit von Desinformation. Auch zur Detektion von digitaler Desinformation setzen die Anbieter auf KI-Lösungen. Ihre technische und organisatorische Entscheidungsmacht wirft Fragen zum Umgang mit den Risiken für die

Grundrechtssphären der Nutzer auf und begründet das Erfordernis, die Voraussetzungen für einen rechtskonformen Einsatz solcher Technologien zur Eindämmung von Desinformation näher zu beleuchten.[1]

Desinformationen können sich negativ auswirken auf den öffentlichen Diskurs, die öffentliche Gesundheit oder Sicherheit, die politische Beteiligung sowie Persönlichkeitsrechte Dritter. Sie können Hate Storms verursachen, Individualrechte verletzen und zu einer irreversiblen Rufschädigung betroffener Personen und Institutionen führen. Strategisch eingesetzte Desinformationen können Hass und Zwietracht säen und darauf zielen, den politischen Gegner zu destabilisieren oder gar Teil hybrider Kriegsführung sein. In normativer Hinsicht ist insbesondere eine polarisierende Wirkung problematisch, weil sie die Kompromissfindung, die existenziell für Demokratien ist, erschwert.[2] Bei einer kontinuierlichen Konfrontation mit Desinformationen und sich widersprechenden Informationen können das Streben nach Wahrheit, Wahrhaftigkeit und Erkenntnisgewinn als Diskursnormen an Bedeutung in der Öffentlichkeit einbüßen.[3] Insbesondere Desinformationen, die Teil (rechts-)extremer Hasspropaganda sind, können radikalisierend wirken und für Anhänger als Grundlage dienen, um reale Gewalt zu legitimieren.[4] Weiterhin können gesundheitsbezogene Desinformationen wesentlich zu einer „Infodemie" in Bezug auf Covid-19 beitragen,[5] zu Selbstschädigungen infolge der Einnahme eines vermeintlichen „Wundermittels" führen oder bewirken, dass der Gebrauch bewährter, evidenzbasierter medizinischer Methoden, Therapien und Medikamente abgelehnt wird und dadurch eine Schädigung der Gesundheit oder sogar lebensbedrohliche Folgen eintreten[6]. Auch können Desinformationen die Glaubwürdigkeit der Wissenschaft untergraben, wenn entgegen eines sehr breiten wissenschaftlichen Konsenses öffentlichkeitswirk-

1 Teile dieses Beitrags wurden bereits in einer Kurzfassung veröffentlicht im Blogbeitrag *Löber*, KI-Lösungen gegen Desinformation in Social Networks – Fragen des grundrechtskonformen und transparenten Einsatzes.

2 *Stark/Stegmann*, Are Algorithms a Threat to Democracy?, 2020, 15.

3 *Jaster/Lanius*, in: Hohlfeld u. a. (Hrsg.), Fake News und Desinformation, 2020, 245 (260) m.w.N.; *Kajewski*, ZfP 2017, 454 (454f.).

4 Näher etwa *Ipsen* u. a., Bericht Rechtsextremismus im Netz 2020/21, 2021, 10 ff. Ein besonders dramatisches Beispiel ist der Anschlag von Hanau im Februar 2020, bei dem zehn Menschen von einem zutiefst rassistischen Täter ermordet wurden, der Verschwörungstheorien in Social Networks verbreitete, die auch Teil des QAnon-Verschwörungsnarrativs sind. S. dazu *Huesmann*, RND vom 11.4.2020.

5 Näher *Islam* u. a., Am. J. Trop. Med. Hyg., 103(4), 2020, 1621 (1621 ff.).

6 Näher *Feldwisch-Drentrup/Kuhrt*, Schlechte und gefährliche Gesundheitsinformationen, 2019, 8 ff.

sam gegenteilige Behauptungen als Auswuchs einer grundlegenden antiwissenschaftlichen und antimedizinischen Haltung aufgestellt werden, z.B. in Bezug auf den menschengemachten Klimawandel[7] oder hinsichtlich Krankheiten und ihrer Ursachen.[8]

Für die wirksame Eindämmung von digitaler Desinformation als sehr vielschichtiges Problem mit einer ganzen Reihe tiefergehender Ursachen und Wirkungen ist die Einbeziehung einer Vielzahl von Akteuren auf diversen gesellschaftlichen, politischen und rechtlichen Ebenen unerlässlich.[9] Einen großen Beitrag zur Adressierung von Teilen dieses Problems können die Anbieter von Social Networks leisten, deren Kommunikationsplattformen zu den wichtigsten Verbreitungskanälen von digitalen Desinformationen, Verschwörungsmythen und Hassreden zählen.[10] Sie haben, anders als Außenstehende, die Möglichkeit, auch koordinierte Desinformationskampagnen aufzudecken und die Verbreitung schädlicher Inhalte mittels algorithmischer Verfahren zu reduzieren. Sowohl die schiere Masse von Beiträgen als auch die Geschwindigkeit, mit der sich Desinformationen und (andere) rechtswidrige Inhalte über das Internet und speziell Social Networks verbreiten, bedingen den Bedarf, automatisierte Erkennungs- und Filtersysteme zu entwickeln und einzusetzen. Auch für die Aufdeckung technisch fortgeschrittener Manipulationsmöglichkeiten, wie Social Bots und Deepfakes, sind technische Lösungen unverzichtbar.[11]

Diese sehr nützlichen, aber auch risikobehafteten Werkzeuge halten in erster Linie die Anbieter der globalen, privaten Online-Plattformen fest in ihren Händen. Sie wenden KI-Technologien zum einen an, um einer rechtlichen Pflicht zur Unterbindung von Rechtsverletzungen nachzukommen, und zum anderen als privatautonome Maßnahme, um plattformeigene Hausregeln („Community Richtlinien", „Gemeinschaftsstandards" etc.) durchzusetzen und das Netzwerk für Nutzende und Werbekunden attraktiv zu halten. Beispielsweise gibt YouTube an, dass von Oktober bis Dezember 2021 99,5 Prozent der entfernten Kommentare von automatischen

7 Vgl. dazu *Jaster/Lanius*, in: Hohlfeld u. a. (Hrsg.), Fake News und Desinformation, 2020, 245 (259f.).

8 S. zum Bestreiten von Krankheiten und ihren Ursachen Seth Kalichman in *Kringiel*, Spiegel Online vom 16.2.2021.

9 *Löber/Roßnagel*, in: Steinebach u. a. (Hrsg.), Desinformation aufdecken und bekämpfen, 2020, 149 (187).

10 *Lazer* u. a., Science 359 (2018), 1094 (1095); *Vosoughi* u. a., Science 359 (2018), 1146 (1146 ff.).

11 Zu technischen Detektionsmöglichkeiten von Desinformation s. etwa *Halvani* u. a., in: Steinebach u. a. (Hrsg.), Desinformation aufdecken und bekämpfen, 2020, 101 ff.

Meldesystemen erkannt wurden.[12] Zu unterscheiden ist zwischen vollautomatisierten und teilautomatisierten Verfahren: Während bei vollautomatisierten Verfahren die Erkennung, Entfernung oder das Downranking von Desinformation in Texten und Bildern sowie von Deepfakes und Social Bots ohne menschliche Beteiligung erfolgen, übernehmen beim teilautomatisierten Einsatz Menschen die einzelfallbezogene Interpretation und Überprüfung der von technischen Systemen erzeugten Meldungen.

2 Technikimmanente Erkenntnisgrenzen und Risiken der KI-Systeme

Der Filtereinsatz zur Eindämmung von Desinformation birgt jedoch einige Konfliktpotenziale, von denen vorliegend die Problematik falsch-positiver Treffer und die mit der enormen technisch-organisatorischen Gestaltungs- und Entscheidungsmacht der Social Networks einhergehenden Risiken fokussiert werden.

2.1 Problematik der Fehltreffer

Trotz immer leistungsfähigerer Systeme kann das Risiko falsch-positiver Treffer in der Regel nicht vollständig ausgeschlossen werden. Fehltreffer können beispielsweise von der Meinungs- und Kunstfreiheit geschützte satirische Darstellungen sein, deren Entfernung weder nach Gesetzesrecht noch nach den privatautonomen Regeln der Anbieter gerechtfertigt ist. Die technischen Systeme können nicht wie Menschen unter Berücksichtigung des Gesamtzusammenhangs der Äußerung zwischen einer Tatsachenbehauptung und einer Meinung unterscheiden, geschweige denn beurteilen, ob eine Tatsachenbehauptung und eine Meinung im engeren Sinne sinnstiftend miteinander verbunden sind, sodass die Äußerung insgesamt als geschützte Meinungsäußerung anzusehen ist.[13] Aufgrund technikimmanenter Erkenntnisgrenzen können sie auch nicht wie Menschen den Wahrheitsgehalt einer Tatsachenbehauptung prüfen oder eine Abwägung der widerstreitenden Grundrechtspositionen leisten. Entsprechendes gilt für das in der Abwägung anzusetzende Gewicht der Meinungsfreiheit, das umso höher ist, je mehr es sich um einen Beitrag zur öffentlichen

12 S. *Google*, YouTube-Community-Richtlinien und ihre Anwendung, 2022.
13 S. zur ständ. Rspr. nur BVerfGE 90, 241 (248); 61, 1 (9); 85, 1 (15f.).

Meinungsbildung handelt, und umso geringer, je mehr es lediglich um emotionalisierende Stimmungsmache gegen einzelne Personen geht.[14]

Wie zuverlässig die Anbieter von Social Networks bestimmte Inhalte und Manipulationen detektieren und welche Werte sie für die automatisierte Entfernung oder die Einordnung als potenzielle Desinformation oder Hassrede voraussetzen, ist nicht bekannt. Im Unterschied zu externen Forschenden verfügen sie über eine riesige Menge an Trainingsdaten sowie über große Personalressourcen und die Mittel, Innovationen zu tätigen, sodass von höheren Trefferquoten als in der externen Forschung auszugehen ist. Die riesigen Datenpools, mit denen lernfähige Systeme für ganz unterschiedliche Einsatzzwecke (z. B. Erkennung von Desinformation und Hassrede, aber auch zur Erstellung von Persönlichkeitsprofilen und für individualisierte Werbung) trainiert und evaluiert werden, liegen in der Hand der globalen, privaten Online-Plattformen, die ihren Datenschatz nur für eigene Zwecke nutzen und bei denen ausreichende Schutzvorkehrungen für betroffene Personen zur Kontrolle über ihre personenbezogenen Daten fehlen.[15]

2.2 Technisch-organisatorische Gestaltungs- und Entscheidungsmacht

Darüber hinaus nimmt die Informationssteuerung und -bündelung durch algorithmische Entscheidungsfindung wirkmächtiger Social Networks erheblichen Einfluss darauf, welche meinungsbildungsrelevanten Inhalte wie wahrgenommen werden. Eine gewisse Diskursstrukturierung wird von Nutzenden aufgrund der Informationsflut im Internet erwartet. Jedoch liegt ein erhebliches Risiko darin, dass Technik im Allgemeinen und KI-Systeme im Speziellen eine besonders große Macht kennzeichnet, die nicht selten verkannt wird und sich auch daraus speist, dass ihr der Machtfaktor nicht anzusehen ist, da sie sehr neutral wirkt.[16] Die Technik des Internets hat die Verwirklichungsbedingungen von (Kommunikations-)Grundrechten und Demokratie verändert – in einiger Hinsicht konnten und können sie verbessert werden, in anderer Hinsicht sind sie neuen internet- und anbieterspezifischen Gefährdungspotenzialen ausgesetzt.[17] So entscheiden die Anbieter mit den von ihnen ausgerichteten Algorithmen und den

14 BVerfG, NJW 2020, 2622 Rn. 29 m.w.N.; BVerfG, GRUR-RS 2021, 44392 Rn. 31.
15 S. zu dieser Problematik *Roßnagel*, Datenspenden für KI – Vertrauen nur mit Grundrechtsvorsorge.
16 *Roßnagel*, MMR 2020, 222 (222) m.w.N.
17 Vgl. *Roßnagel*, MMR 2020, 222 (225).

für die Nutzung aufgestellten Bedingungen, Regeln sowie deren Vollzug über Verwirklichungsbedingungen von Grundrechten im digitalen Raum und folglich über Grundlagen individueller und gesellschaftlicher Freiheit.[18] Durch ihre technisch-organisatorische Gestaltungs- und Entscheidungsmacht können sie die mediale Öffentlichkeit in schädigender oder zumindest nicht transparenter Weise strukturieren.

Seit einiger Zeit steht der – nunmehr von der Whistleblowerin Frances Haugen gestützte – Vorwurf im Raum, Facebook setze prioritär auf Wachstum und Werbeeinnahmen und nehme dafür bewusst in Kauf, Manipulationsversuche nicht hinreichend zu bekämpfen und Algorithmen einzusetzen, die spalterische und schädliche Inhalte fördern, da sie besonders viele Nutzerreaktionen verursachen.[19] Konkret beschuldigen Vertriebene der Rohingya aus Myanmar Facebook in beispiellosen Sammelklagen auf Schadensersatz von rund 150 Milliarden Dollar. Facebooks Algorithmen und die unterlassene Eindämmung von Hetze und Desinformation sollen reale Gewalt gegen sie angefacht haben.[20]

Die ohnehin schon eingeläutete Phase der stärkeren Regulierung der Internetgiganten erhielt durch diese Enthüllungen nochmals Aufwind. Dies gilt nicht nur speziell mit Blick auf digitale Desinformation und Hetze, sondern ebenfalls für die Rahmenbedingungen, Algorithmen und KI. Denn klar dürfte mittlerweile jedem sein: Wer die Macht über die technische, organisatorische und inhaltliche Gestaltung der Social Networks – und damit über zentrale Kommunikationskanäle unserer heutigen Zeit – hat, trägt eine enorme gesellschaftliche Verantwortung für das friedliche Miteinander und die Funktionsfähigkeit demokratischer Abläufe.[21]

3 KI-Einsatz in der Internet(selbst)regulierung als Balanceakt

Um diese Risiken für die Grundrechtssphären der Nutzenden möglichst effizient zu reduzieren sowie Meinungspluralität und den unverfälschten Meinungsbildungsprozess als Grundpfeiler der Demokratie zu schützen, bedarf es mehrerer, ineinandergreifender und sich ergänzender Lösungsan-

18 *Roßnagel*, MMR 2020, 222 (223).

19 S. etwa *Tagesschau* vom 5.10.2021.

20 In Myanmar stellt sich die Internetnutzung und das Informationsrepertoire der Menschen anders dar als z. B. in Deutschland. Es soll dort gleich viele Internet- und Facebook-Nutzer geben – Facebook ist wohl für viele Menschen dort die Hauptinformationsquelle. S. *Kreye*, Süddeutsche Zeitung vom 7.12.2021.

21 *Kühling*, ZUM 2021, 461 (461).

sätze, die eine rechtliche Querschnittsmaterie betreffen. Dabei muss die Internet(selbst)regulierung, die sich zunehmend des Einsatzes von KI bedient, den schwierigen Balanceakt vollziehen, dass der Einsatz nicht selbst eine Rechtsverletzung birgt (etwa Verstoß gegen das datenschutzrechtliche Verbot automatisierter Einzelentscheidungen) oder herbeiführt (etwa Entfernung eines rechtmäßigen, von Art. 5 Abs. 1 GG geschützten Beitrags) und zudem Nutzende nicht in Unkenntnis darüber gelassen werden, dass und in welcher Art und Weise sie automatisierten Entscheidungen ausgesetzt sind. Den Online-Plattformen auferlegte „Filterpflichten" sowie rechtliche Grenzen des KI-Einsatzes verfassungsrechtlicher, medienrechtlicher und datenschutzrechtlicher Natur und Transparenzverpflichtungen etwa aus dem Medienstaatsvertrag (MStV) und dem Netzwerkdurchsetzungsgesetz (NetzDG) adressieren diese Herausforderungen mittelbar und unmittelbar.

Der (technische) Umgang mit Desinformationen ist gesetzlich nicht spezifisch geregelt. Denn die freiheitlich-demokratische Grundordnung vertraut darauf, dass im Prozess der Meinungsbildung und geistigen Auseinandersetzung auch drastische, extreme Positionen sowie Falschinformationen durch Gegenrede relativiert werden,[22] sodass sich im Ergebnis die „Macht der Vernunft" und die vernünftigste Meinung durchzusetzen vermag.[23] Der Meinungskampf und die Bildung von Willensentscheidungen werden verstanden als „process of trial and error", der „nicht immer objektiv richtige Ergebnisse" liefert, aber doch „durch die ständige gegenseitige Kontrolle und Kritik die beste Gewähr für eine (relativ) richtige politische Linie als Resultante und Ausgleich zwischen den im Staat wirksamen politischen Kräften gibt".[24] Diese Verfassungserwartung zeigt, dass aus den Kommunikationsgrundrechten keine generelle „Wahrheitspflicht" folgen kann.[25] Jenseits rechtswidriger Äußerungen vor allem aus den Bereichen des Strafrechts und Persönlichkeitsrechts sind die Aufgaben des Staates daher sehr begrenzt und richten sich maßgeblich auf die Aufrechterhaltung der Bedingungen für die Selbstorganisation gesellschaftlicher Kommunikation.[26] Dem Grundsatz der staatlichen Zurückhaltung folgend greift die

22 Etwa *Klein*, in: Dürig u. a. (Hrsg.), Grundgesetz-Kommentar, 2020, Art. 41 GG Rn. 123.
23 Vgl. *Kloepfer*, in: Isensee/Kirchhof (Hrsg.), HStR III, 2005, § 42, Rn. 14.
24 BVerfGE 5, 85 (135); 69, 315 (345f.).
25 *Jestaedt*, in: Merten/Papier (Hrsg.), 2011, § 102, Rn. 36; *Degenhart*, in: Bonner Kommentar zum GG, 2021, Art. 5 Abs. 1 und 2 Rn. 118.
26 *Löber/Roßnagel*, in: Steinebach u. a. (Hrsg.), Desinformation aufdecken und bekämpfen, 2020, 149 (187).

Regulierung erst dort ein, wo die Kräfte der gesellschaftlichen Auseinandersetzung nicht ausreichen, und soweit es darum geht, notwendige Rahmenbedingungen, u. a. für einen vernünftigen, transparenten Umgang mit den KI-Systemen und der Content-Moderation mächtiger Online-Plattformen im digitalen Raum, zu schaffen.

3.1 Verpflichtender Einsatz bei duplizierten und sinngleichen Inhalten

Diese Herausforderungen und Lösungsansätze zeigen sich beispielsweise im Rahmen der Rechtsprechung und Debatte zur Auferlegung von Pflichten an Diensteanbieter wie Facebook, zukunftsgerichtet nicht nur die Weiterverbreitung eines konkreten rechtswidrigen Inhalts, etwa einer verleumderischen Falschbehauptung, zu verhindern, sondern auch Duplikate und sogar sinngleiche Beiträge anderer Nutzer (weltweit) zu entfernen.[27] Schließlich können die technischen Erkennungssysteme verhindern, dass Betroffene gegen jeden einzelnen duplizierten und ähnlich-duplizierten Beitrag separat vorgehen müssen – bei Desinformationskampagnen und Shit Storms angesichts der Funktionslogiken im digitalen Raum ein oftmals aussichtsloses Unterfangen. Die beeinträchtigende Wirkung einer Äußerung ist gesteigert, wenn sie in wiederholender und anprangernder Weise sowie besonders sichtbar im Internet als verstärkendem Medium unter Berücksichtigung der konkreten Breitenwirkung getätigt wird.[28]

In der nationalen Rechtsprechung legt der BGH Diensteanbietern bei Hinweisen auf klare Rechtsverletzungen bereits seit einigen Jahren Filterpflichten auf. Danach kann eine Verpflichtung zur notwendigen, automatischen Kontrolle aller Inhalte, z. B. mit Wortfiltern, um Duplikate und ähnlich klare Rechtsverletzungen zu unterbinden, zumutbar sein, wobei eine manuelle Nachkontrolle auf bestimmte, vorab gefilterte Inhalte beschränkt sein muss.[29] Mit dieser Rechtsprechung wurden dem Wettbewerbs-, Urheber- und Markenrecht entspringende Unterlassungsansprüche auf im Kern gleichartige Verletzungshandlungen erstreckt.[30] Die Übertrag-

27 Vgl. EuGH, Urt. v. 3.10.2019 – C-18/18, ECLI:EU:C:2019:821 – Glawischnig-Piesczek/Facebook.

28 BVerfG, NJW 2020, 2622 Rn. 34; NJW 2020, 300 Rn. 125.

29 BGH, ZUM 2013, 874 Rn. 61.

30 Vgl. zu manuellen Entfernungen gleichartiger Verstöße BGH, NJW 2019, 1142 Rn. 18 ff.; BGH, MMR 2014, 190 Rn. 18; BGH, MMR 2011, 385 Rn. 26; BGH, NJW-RR 2006, 1048 Rn. 36.

barkeit auf Sachverhalte aus dem Äußerungsrecht ist jedoch noch nicht geklärt.[31]

Indessen hat der EuGH vor dem Hintergrund der Auslegung des Verbots allgemeiner Überwachungspflichten für Host-Provider gemäß Art. 15 E-Commerce-RL entschieden, dass im Nachgang zu einer Rechtsverletzung Diensteanbieter grundsätzlich verpflichtet werden können, auch wort- und sinngleiche Inhalte zu verhindern. Eine übermäßige, Art. 15 E-Commerce-RL zuwiderlaufende Verpflichtung könne dadurch verhindert werden, dass die Überwachung und Nachforschung „auf die Informationen beschränkt sind, die die in der Verfügung genau bezeichneten Einzelheiten enthalten" und dass „ihr diffamierender Inhalt sinngleicher Art den Hosting-Anbieter nicht verpflichtet, eine autonome Beurteilung vorzunehmen, so dass er auf automatisierte Techniken und Mittel zur Nachforschung zurückgreifen kann".[32] Folglich verläuft nach der Judikatur des EuGH die Grenze der Verpflichtung dort, wo der Anbieter zu einer autonomen Entscheidung gezwungen wäre, auch um die unternehmerische Freiheit hinreichend zu berücksichtigen. Allerdings war in dem betreffenden Fall gerade keine komplexe Interessenabwägung erforderlich, da es sich um eindeutige Beleidigungen handelte.[33] Daher blieb offen, wie weit die Verhinderungspflichten in weniger eindeutigen äußerungsrechtlichen Fällen reichen und inwieweit das Risiko von Fehltreffern hingenommen oder durch manuelle Nachkontrollen abgefedert werden müsste. Dabei steht die Rechtsprechung vor der Herausforderung, dass sie kaum abschätzen kann, wie zuverlässig die technischen Erkennungssysteme der Anbieter tatsächlich funktionieren und inwieweit in konkreten Fällen die Verletzung von Kommunikationsgrundrechten durch eine automatisierte Filterung droht.

3.2 Reichweite der Befugnisse bei privatautonomen Maßnahmen

Auch die Anbieter von Social Networks haben bei der Content-Moderation einen äußerst schwierigen Balanceakt auszuführen. Erwecken sie den Eindruck, zu viel zu entfernen, sind sie dem Vorwurf ausgesetzt, den für

31 Offen gelassen BGH, NJW 2019, 1142 Rn. 18 ff. zur Wortberichterstattung; wohl bejahend in BGH, ZUM-RD 2019, 203 Rn. 44. Vgl. auch *Specht-Riemenschneider*, MMR 2019, 801 (801f.); LG Würzburg, MMR 2017, 347 (349).

32 EuGH, Urt. v. 3.10.2019 – C-18/18, ECLI:EU:C:2019:821, Rn. 46 – Glawischnig-Piesczek/Facebook.

33 Ebenso *Spindler*, NJW 2019, 3274 (3275).

die öffentliche Meinungsbildung essenziellen „Kampf der Meinungen" nicht zuzulassen, Kommunikationsgrundrechte nicht zu achten und sich vertragswidrig gegenüber betreffenden Nutzenden zu verhalten. Tun sie zu wenig, werden sie kritisiert, Hass, Hetze und Desinformation zu dulden sowie zur Verfälschung des Meinungsbildungsprozesses beizutragen und riskieren eine Inanspruchnahme etwa auf Schadensersatz, Unterlassung sowie eine mögliche strafrechtliche Verantwortlichkeit.[34] Im Rahmen ihrer weitreichenden privatautonomen Befugnisse können sie die Online-Plattform grundsätzlich unter Achtung der allgemeinen Gesetze frei gestalten und freiwillig auch über bestehende gesetzliche Regeln hinaus Sanktionen für bestimmte Inhalte, Manipulationen und Formen von Desinformationen vorsehen.[35] Die in ihren Nutzungsbedingungen aufgestellten Regeln dürfen sich nach umstrittener, aber überzeugender Auffassung auf nur AGB-widrige – und nicht zugleich gesetzeswidrige – Inhalte erstrecken, wobei sie die mittelbaren Grundrechtsgefährdungen, namentlich die Kommunikationsgrundrechte der Inhalteersteller, die Grundrechtssphäre der von einer versagten Rezeption und Anschlusskommunikationen betroffenen Nutzenden, das allgemeine Persönlichkeitsrecht sowie das Gleichbehandlungsgebot, hinreichend berücksichtigen müssen.[36]

In jedem Fall müssen Maßnahmen wie die automatisierte und nichtautomatisierte Entfernung von Beiträgen oder die Sperrung von Nutzerkonten auf Basis klarer, vorab formulierter Kriterien sowie zumutbarer Anstrengungen zur Aufklärung des Sachverhalts, zu denen auch verfahrensrechtliche Absicherungen gehören, erfolgen.[37] Mit zunehmender Macht, gesellschaftlicher Relevanz und Einflussnahme auf die Kommunikationsprozesse steigt das Ausmaß der rechtlichen Verpflichtung der Anbieter.[38] Je größer, wirkmächtiger, „unausweichlicher" die Online-Plattformen sind, desto höhere Anforderungen sind auch an die Zuverlässigkeit ihrer automatisierten Systeme und Entscheidungen sowie die vorzusehenden technischen und rechtlichen Schutzmaßnahmen zu stellen. Staatliche Befugnisse sind hier, jedenfalls soweit es nicht um nachweisliche strukturelle Meinungsvielfaltsgefährdungen durch Filtersysteme geht, vornehmlich

34 Vgl. BGH, NJW 2021, 3179 Rn. 77.
35 *Löber/Roßnagel*, MMR 2019, 71 (75); *Dreyer* u. a., Desinformation, 2021, 45.
36 Bestätigt von BGH, NJW 2021, 3179 Rn. 58 ff. m.w.N.; *Ingold*, in: Unger/v. Ungern-Sternberg (Hrsg.), Demokratie und Künstliche Intelligenz, 2019, 183 (198).
37 Vgl. BGH, NJW 2021, 3179 Rn. 79 ff.; BVerfG, NJW 2018, 1667 Rn. 46 ff.
38 Vgl. BVerfGE 128, 226 (248 ff.); verstärkt durch BVerfG, NJW 2015, 2485 (2486); s. auch *Löber/Roßnagel*, in: Steinebach u. a. (Hrsg.), Desinformation aufdecken und bekämpfen, 2020, 149 (173).

darauf gerichtet, diese privatautonomen Verfahren gesetzlich zu rahmen und die Beachtung grundrechtssichernder Mindeststandards im Sinne der Kommunikationsgrundrechte sicherzustellen.[39]

3.3 Technische und rechtliche Mechanismen zum Schutz der Kommunikationsgrundrechte

Durch den Einsatz von Filtertechnologien zur Eindämmung von Desinformation darf es nicht zu unverhältnismäßigen Grundrechtseingriffen kommen. Wichtige Schutzmechanismen stellen insbesondere ein wirksames Zusammenspiel von automatisierter Filterung und manueller Kontrolle sowie ein verfahrensbasierter Grundrechtsschutz dar. So ermöglicht die Parametrisierung der Erkennungs- und Filtersysteme verschiedene Reaktionsmöglichkeiten, die nicht lediglich ein binäres System darstellen, sondern abhängig von Übereinstimmungs- bzw. Wahrscheinlichkeitswerten gestufte Maßnahmen und sowohl vollautomatisierte als auch teilautomatisierte Reaktionen in verschiedenen Ausprägungen umfassen. Auf diese Weise können z. B. vollautomatisierte Filterungen nur bei eindeutigen, äußerst hohen Übereinstimmungswerten ausgeführt und bei weniger eindeutigen Ergebnissen die menschliche Überprüfung durch geschultes Personal sowie die Möglichkeit zur Stellungnahme für die beteiligten Personen initiiert werden.[40] Bereits die Annäherung an richtige Ergebnisse in solchen teilautomatisierten Verfahren kann einen erheblichen Mehrwert darstellen, da die Auffindbarkeit der desinformativen Inhalte und Manipulationen ermöglicht oder zumindest beschleunigt wird und Fehltreffer im manuellen Überprüfungsprozess ausgeschlossen werden können.[41] Außerdem kann das KI-System den Mitarbeitenden als Entscheidungshilfe ähnliche Fälle und deren rechtliche Bewertung anzeigen sowie eine Vorsortierung nach Mustern vornehmen, um den Nutzen des Vorfilterns zu optimieren. Auch ein stärkeres Interagieren von Mensch und KI, bei dem Menschen die Fehler der KI direkt an diese zurückspiegeln, kann Falscherkennungen minimieren.

Ein verfahrensbasierter Grundrechtsschutz ist aufgrund der mittelbaren Drittwirkung des Gleichheitssatzes und der Kommunikationsfreihei-

---

39 So auch *Dreyer* u. a., Desinformation, 2021, 45; *Ingold*, in: Unger/v. Ungern-Sternberg (Hrsg.), Demokratie und Künstliche Intelligenz, 2019, 183 (201).

40 S. *Raue/Steinebach*, ZUM 2020, 355 (363) in Bezug auf Upload-Filter.

41 Vgl. *Kastl*, GRUR 2016, 671 (673).

ten auch bei privatautonomen Maßnahmen der Social Networks notwendig. Er erfordert, dass niedrigschwellige plattforminterne Verfahren und Möglichkeiten externer außergerichtlicher Streitschlichtung für Nutzende verfügbar sind, um eine hinreichend bestimmte Tatsachengrundlage der Entscheidungen sowie ein effektives Vorgehen gegen fehlerhafte Maßnahmen zu gewährleisten.[42] Geht es um die Entfernung äußerungsbezogener Inhalte, gehört zu den Organisations- und Verfahrensvorschriften auch die Gelegenheit zur Stellungnahme der Inhaltersteller vor Ergreifung der Maßnahme – jedenfalls abseits sehr eindeutiger Fälle und sehr hoher Trefferquoten – sowie die Begründung von Entscheidungen, um insbesondere willkürliche Maßnahmen auszuschließen, transparent zu handeln und eine zumutbare Aufklärung des Sachverhalts durch Menschen vorzunehmen.[43]

Angesichts dieser technischen und rechtlichen Möglichkeiten und Anforderungen, die von Fehltreffern ausgehenden Gefahren für die Grundrechtssphären der Nutzenden effizient zu reduzieren, führt das Risiko falsch-positiver Treffer auch im sehr grundrechtssensiblen Äußerungsrecht jedenfalls nicht per se zum Ausschluss solcher Filtertechnologien. Jedoch ist die Möglichkeit, vollautomatisierte Verfahren rechtsverträglich einzusetzen, hier sehr stark eingeschränkt und wohl nur bei äußerst treffsicheren Ergebnissen, etwa bei Bildern oder Duplikaten eindeutig rechtswidriger Desinformation, in Betracht zu ziehen, um Overblocking zu vermeiden.

4 Transparenzvorgaben gegen weitreichende Wissensasymmetrien

In Anbetracht des weitreichenden Wissensvorsprungs der Diensteanbieter hinsichtlich der von ihnen eingesetzten algorithmischen Verfahren sind Transparenzvorgaben ein zentraler Pfeiler der Regulierung. Transparenz im Umgang mit Desinformationen ist besonders bedeutsam, um nachvollziehen zu können, in welcher Weise und nach welchen Kriterien die Anbieter auf den freien Diskurs einwirken und diesen vor Manipulationen zu schützen versuchen. Dabei geht es neben Sanktionen, wie das Löschen und Sperren von Inhalten und Nutzerkonten, um weitere Formen der Content-Moderation, wie die Kennzeichnung oder das – mitunter für Nutzende nicht erkennbare – Downranking von Falschinformationen, bei denen die Verletzung von Kommunikationsgrundrechten ebenfalls nicht

42 BGH, NJW 2021, 3179 Rn. 79 ff.; *Reinhardt/Yazicioglu*, DSRITB 2020, 819 (826).
43 Vgl. BVerfG, NJW 2018, 1667 Rn. 46.

ausgeschlossen ist, und die derzeit nicht spezifisch gesetzlich reguliert werden. Freiwillig zeigen Social Networks indes nur sehr begrenzt Transparenz. In der Regel beschränken sie sich auf sehr allgemeine, vage Angaben zum Einsatz von Erkennungs- und Filtersystemen in einem bestimmten Zeitraum und nicht in Bezug auf konkrete Beiträge oder auf Deutschland.[44] Solche freiwilligen Transparenzangaben oder Transparenzpflichten ohne hinreichende Überprüfbarkeit stehen unter Vergeblichkeitsverdacht, da sie darauf ausgerichtet sein können, einer bestimmten, unternehmensfreundlichen Erzählrichtung zu folgen.[45]

4.1 Netzwerkdurchsetzungsgesetz: Informationen zu eingesetzten Verfahren zur automatisierten Erkennung von Inhalten

Der nationale Gesetzgeber versucht den erheblichen Wissensasymmetrien mit spezifischen Transparenzpflichten im Netzwerkdurchsetzungsgesetz und Medienstaatsvertrag abzuhelfen. Im Netzwerkdurchsetzungsgesetz sind keine spezifischen (Transparenz-)Pflichten im Umgang mit Desinformation und deren automatisierter Aufdeckung enthalten. Im Hinblick auf die von Anbietern von Social Networks vorzuhaltenden Beschwerdemanagementsysteme zur Entfernung bestimmter rechtswidriger Inhalte innerhalb vorgegebener Fristen sind auch Desinformationen erfasst, soweit sie Straftatbestände wie Verleumdung nach § 187 StGB oder Holocaust-Leugnung nach § 130 Abs. 3 StGB erfüllen (vgl. den Katalog in § 1 Abs. 3 NetzDG).[46] Außerdem sind die Anbieter gemäß § 2 Abs. 2 Nr. 2 NetzDG verpflichtet, allgemeine Angaben zu eingesetzten Verfahren zur automatisierten Erkennung von Inhalten, die wegen ihrer gesetzlichen oder vertraglichen Unzulässigkeit entfernt werden sollen, mitzuteilen. Diese Vorgaben sollen der Erkenntnis Rechnung tragen, „dass erhebliche Fortschritte beim automatisierten Aufspüren von entsprechend unzulässigen Inhalten gemacht worden sind und die Öffentlichkeit darüber an zentraler Stelle informiert werden sollte".[47] Aus dem Gesetzeswortlaut („automatisierte Er-

44 S. z. B. die Angaben von *Google*, YouTube-Community-Richtlinien und ihre Anwendung, 2022.

45 Vgl. *Cornils*, ZUM 2019, 89 (102), in Bezug auf die Regelungen für Medienintermediäre im MStV. S. die Enthüllungen des ehemaligen Vize-Marketing-Chefs von Facebook Boland in *Alba/Mac*, New York Times vom 20.8.21.

46 S. zu den Verpflichtungen aus dem NetzDG auch *Löber/Roßnagel*, in: Steinebach u. a. (Hrsg.), Desinformation aufdecken und bekämpfen, 2020, 149 (168 ff.).

47 BT-Drs. 19/18792, 42.

kennung", „Überprüfung der Ergebnisse der automatisierten Verfahren durch den Anbieter") folgt, dass die Transparenzangaben sowohl für vollautomatisierte als auch für teilautomatisierte Löschungen oder Sperrungen zu erbringen sind.[48] Dass die Regelung nicht nur Inhalte umfasst, die gesetzeswidrig sind, sondern auch solche, die gegen vertragliche Bestimmungen im privatrechtlichen Verhältnis von Anbietern und Nutzenden verstoßen, ist essenziell, da ein Großteil der Maßnahmen der Anbieter der Durchsetzung ihrer Nutzungsbedingungen dient. Beispielsweise hat Facebook bereits im Gesetzgebungsverfahren klargestellt, dass es „Technologie zur automatischen Erkennung von Inhalten einsetzt, die den Gemeinschaftsstandards von Facebook widersprechen", nicht jedoch für eine Prüfung, ob sie deutsches Recht verletzen.[49]

Obgleich die Regelung den Umgang mit Desinformationen nicht explizit aufgreift, gehören automatisierte Verfahren zur Erkennung von Desinformationen, die entfernt werden sollen, zu den Verfahren gemäß § 2 Abs. 2 Nr. 2 NetzDG. Indes ist angesichts der vagen Bestimmungen fraglich, inwieweit der Vorschrift eine Pflicht zu entnehmen ist, nähere Informationen zu diesen Verfahren vorzulegen. Wegen der Begrenzung auf allgemein gehaltene Informationen steht nicht zu erwarten, dass sie eine Überprüfung zuließen, ob und inwieweit sie tatsächlich rechtskonform gestaltet und eingesetzt werden. Als echte Kontrollmaßstäbe sind diese Vorgaben kaum geeignet. Zumindest können die Vorgaben daneben als sanfter Anreiz verstanden werden, die ohnehin eingesetzten technischen Systeme besonders sorgsam und grundrechtsschonend einzusetzen. Dafür spricht beispielsweise, dass die Anbieter über qualitätssichernde Maßnahmen sowie etwaige Überprüfungen von Filterergebnissen durch Menschen berichten müssen. Diese Mensch-Maschine-Interaktionen sind besonders relevant für die Ausfilterung falsch-positiver Ergebnisse und den Grundrechtsschutz der Nutzenden.

4.2 Medienstaatsvertrag: grobe Einblicke in Sortier-, Priorisierungs- und Selektierungsmethoden und Diskriminierungsverbot für journalistisch-redaktionell gestaltete Inhalte

Auch die Transparenzvorgaben für Medienintermediäre gemäß § 93 Abs. 1, 3 MStV sind sehr weiche Verpflichtungen. Sie sollen dem Schutz

48 S. auch BT-Drs. 19/18792, 42.
49 S. *Facebook*, Stellungnahme NetzDG-E, 2020, 5.

der Meinungs-, Angebots- und Anbietervielfalt dienen und verlangen allgemein gehaltene Angaben zu Kriterien u. a. über die Selektion und Präsentation von Inhalten sowie die Funktionsweise der eingesetzten Algorithmen. Weder verpflichten sie dazu, Nutzende darüber zu informieren, welche Kriterien ex-post im konkreten Fall ausschlaggebend für die Anzeige eines Inhaltes waren,[50] noch statuieren sie eine spezifische Pflicht, konkret auf den algorithmusbasierten Umgang mit Desinformation und anderen Falschinformationen einzugehen. Jedoch sind nach § 93 Abs. 1 Nr. 1 und 2 MStV allgemeine Informationen hierzu mitzuteilen, sofern z. B. der Wahrheitsgehalt oder die Seriosität relevante Kriterien für Zugang, Verbleib, Aggregation, Selektion, Präsentation oder Gewichtung von Inhalten sind. Angesichts der publikumsorientierten „Einfachheitskonzeption"[51] der verlangten Angaben müssen die sehr komplexen Algorithmen, die sich mit dem Einsatz von KI ständig verändern und wohl nahezu unüberschaubar sind,[52] in eine sehr einfache Sprache übersetzt werden. Die starken Vereinfachungen bergen die Gefahr von (ungewollten) Verfälschungen durch die Berichtenden und mangelnder Überprüfbarkeit seitens der Aufsicht. Dennoch können die Angaben, jedenfalls soweit sie präzise genug, nicht schwer verständlich und leicht auffindbar sind, ein taugliches Mittel darstellen, zumindest grobe Einblicke in Sortier-, Priorisierungs- und Selektierungsmethoden von Medienintermediären zu erhalten und Nutzende für diese Thematik zu sensibilisieren.

Indes gewährleisten die Transparenzvorgaben alleine freilich keine Meinungs-, Angebots- und Anbieter- sowie Nutzungsvielfalt auf den digitalen Kommunikationsplattformen. Im Übrigen folgt auch aus der strengen mittelbaren Drittwirkung per se keine Neutralitätspflicht der Anbieter, die weiterhin Grundrechtsberechtigte sind.[53] Zudem ist die Entscheidung der Nutzenden für personalisierte Angebote und damit das Risiko, sich in digitalen Echokammern zu bewegen, als Ausdruck ihrer Informationsfreiheit zu berücksichtigen.[54] Dementsprechend findet sich in § 94 MStV lediglich ein Diskriminierungsverbot für journalistisch-redaktionell gestaltete Inhal-

50 A.A. *Schwartmann* u. a., Transparenz bei Medienintermediären, 2020, 133.
51 *Cornils*, ZUM 2019, 89 (102).
52 Kritisch und vor diesem Hintergrund eine Informationspflicht über die Kriterien des Filterns von Nachrichten als „regulatorischen Schlag ins Wasser" ablehnend *Drexl*, ZUM 2017, 529 (537, 541f.); ähnlich *Ladeur/Gostomzyk*, K&R 2018, 686 (690f.).
53 *Ingold*, in: Unger/v. Ungern-Sternberg (Hrsg.), Demokratie und Künstliche Intelligenz, 2019, 183 (200).
54 So auch *Martini*, Blackbox Algorithmus, 2019, 103, 225.

te, auf deren Wahrnehmbarkeit die Medienintermediäre besonders hohen Einfluss haben. Sie dürfen nicht entgegen der nach § 93 Abs. 1 bis 3 MStV zu veröffentlichenden Kriterien ohne sachlichen Grund systematisch benachteiligt werden. Das Diskriminierungsverbot ist mithin auf die Abwesenheit unsachgemäßer, rechtswidriger Einflussnahme und Manipulation hinsichtlich journalistisch-redaktioneller Angebote gerichtet und schreibt nicht im Sinne eines Must-Carry-Regimes vor, bestimmte Inhalte, denen ein gesteigertes öffentliches Interesse zukommt, zu transportieren und (privilegiert) auffindbar zu machen.[55] Bezüglich der Handhabung von Desinformation beinhaltet es keine verpflichtenden und zielgerichteten Vorgaben. Es hindert die Anbieter allerdings nicht daran, desinformative Inhalte auf Grundlage ihrer Nutzungsbedingungen z. B. in der Sichtbarkeit einzuschränken oder zu entfernen, da die Integrität des Dienstes und die jeweiligen von Desinformation ausgehenden Risiken für verschiedene Rechtsgüter sachliche Gründe für eine unterschiedliche Behandlung der Angebote darstellen können.

Gegenwärtig rechtfertigen auch bestehende Diskursfragmentierungen des öffentlichen Meinungsbildungsprozesses es nicht, wirkmächtige Intermediäre entgegen ihres Geschäftsmodells und individueller Interessen der Nutzenden sowie deren Informationsfreiheit zur Privilegierung von Inhalten, denen ein besonderer Mehrwert für die öffentliche Kommunikation zukommen soll, zu verpflichten.[56] Es müsste plausibel begründbar sein, dass die gesamtgesellschaftlich integrierte Rezeption insbesondere von politik- und nachrichtenbezogenen Inhalten signifikant gesunken ist oder dies zu befürchten steht.[57] Jedoch rezipiert der Großteil der Internetnutzenden in Deutschland regelmäßig Nachrichten außerhalb von Social Networks.[58] Die Nutzung von Social Networks ist (noch) weitestgehend

55 Vgl. *Heidtke*, Meinungsbildung und Medienintermediäre, 2020, 341f.; *Zimmer*, ZUM 2019, 126 (129); s. auch *Ladeur/Gostomzyk*, K&R 2018, 686 (691), die insoweit die Frage aufwerfen, ob es sich bei dem Diskriminierungsverbot „nicht vorrangig um einen Diskriminierungsschutz von Massenmedien im Wettbewerb zu anderen Inhalten" handele.

56 A.A. *Mitsch*, DVBl 2019, 811 (817f.); *Schwartmann* u. a., Transparenz bei Medienintermediären, 2020, 158 ff.

57 *Mengden*, Zugangsfreiheit und Aufmerksamkeitsregulierung, 2018, 398f., 402.

58 *Hölig* u. a., Digital News Report, 2021, 5: Nachrichten im linearen Programmfernsehen sind bei Betrachtung der Einzelgattungen der am meisten gewählte Zugangsweg. Nur eine sehr geringe Anzahl der Erwachsenen (vier Prozent) konsumiert ausschließlich über Social Networks Nachrichten.

eingebettet in einen breiten Mix verschiedener Medienkanäle.[59] Nachrichtenformate im linearen Programmfernsehen und insbesondere des öffentlich-rechtlichen Rundfunks werden zurzeit in beachtlichem Umfang wahrgenommen.[60] Deren integrative Wirkung kann Fragmentierungen entgegenwirken.

4.3 Kennzeichnungspflichten von Social Bots und Deepfakes

Ein weiterer bedeutsamer Schritt zur Förderung der Integrität der Kommunikation in Social Networks sowie zum Schutz des individuellen und öffentlichen Meinungsbildungsprozesses war die Einführung der Kennzeichnungspflicht von Social Bots für Anbieter von Social Networks im novellierten Medienstaatsvertrag.[61] Die sehr weich und offen im Sinne einer Bemühenspflicht formulierte Regelung in § 93 Abs. 4 MStV, für die Kennzeichnung „Sorge zu tragen", überlässt es den Anbietern, mit welchen technischen Mitteln sie die Identifizierung durchführen, und ob und inwieweit sie menschliche Entscheider zur Überprüfung der technischen Identifizierung von Bots einbinden. Indem ihnen der nötige technische Handlungsspielraum zur Erkennung und Kennzeichnung der Bots eingeräumt wird, ist die Regelung offen für technische Weiterentwicklungen und flexibel umsetzbar.[62] Da die Markierung der automatisierten Kommunikation ein wesentlich geringerer Eingriff als etwa eine Löschung von Bot-Beiträgen und Bot-Konten darstellt und lediglich auf Herstellung von Transparenz bezüglich der Bot-Eigenschaft zielt, mithin nur eine Nebensächlichkeit der Kommunikationsverbreitung betrifft, ist die Regelung verhältnismäßig und mit der Meinungsfreiheit bzw. der allgemeinen Handlungsfreiheit vereinbar.[63]

Während sich diese Kennzeichnungspflicht spezifisch auf automatisierte Kommunikation mittels einem dem äußeren Erscheinungsbild nach für die Nutzung durch natürliche Personen bereitgestellten Nutzerkonto in Social Networks bezieht (vgl. § 18 Abs. 3 Satz 1 MStV), sieht auf europäi-

59 S. *Hölig* u. a., Digital News Report, 2021, 13 ff.; *Stark/Stegmann*, Are Algorithms a Threat to Democracy?, 2020, 20f.
60 S. *Hölig* u. a., Digital News Report, 2021, 22: 68 Prozent der Befragten konsumieren regelmäßig Nachrichten des öffentlich-rechtlichen Rundfunks.
61 Eingehend *Löber/Roßnagel*, MMR 2019, 493 (493 ff.).
62 *Löber/Roßnagel*, MMR 2019, 493 (498).
63 *Löber/Roßnagel*, MMR 2019, 493 (497); *Dreyer* u. a., Desinformation, 2021, 67.

scher Ebene der Entwurf der KI-Verordnung (KIVO-E)[64] eine umfassendere Transparenzpflicht für Mensch-Maschine-Interaktionen vor. Konkret verpflichtet Art. 52 Abs. 1 Satz 1 KIVO-E Anbieter, KI-Systeme, die für Interaktionen mit natürlichen Personen bestimmt sind, so zu konzipieren, dass die jeweiligen Personen darüber informiert werden, dass es sich um ein KI-System handelt. Ausgenommen von der Informationspflicht sind Konstellationen, in denen die KI-Eigenschaft aufgrund der Umstände und des Kontexts der Nutzung offensichtlich ist. Während etwa bei körperlichen Gegenständen wie Staubsaugrobotern diese Eigenschaft in der Regel offensichtlich sein wird,[65] ist gerade im Bereich digitaler Kommunikation, in der Nutzende ihr Gegenüber nicht oder nur auf einem Bild oder in einem Video sehen können, die Mensch-Maschine-Interaktion nicht von vornherein transparent. Daher fallen auch Chat Bots und Social Bots unter die Kennzeichnungspflicht. Bislang ist aber noch nicht klar geregelt, wie die Information der jeweiligen Person über die KI-Eigenschaft erfolgen muss.[66]

Weiterhin geht der Verordnungsentwurf über den Rechtsrahmen auf Bundesebene hinaus, indem Art. 52 Abs. 3 KIVO-E eine Kennzeichnungspflicht für Deepfakes statuiert, die sich an die Nutzer von KI-Systemen richtet. Da mittels Deepfakes in besonders tiefgreifender Weise andere Personen diskreditiert und manipulativ erfundene oder veränderte Botschaften mit erheblichen Bedrohungspotenzialen für demokratische Prozesse kreiert werden können, wäre eine schärfere Regelung, z. B. ein Verbot politischer Deepfakes in dem Zeitraum vor einer politischen Wahl, nicht fernliegend gewesen.[67] Allerdings ist die Kennzeichnung von Deepfakes ausreichend und notwendig, um Rezipierende mit dem Wissen auszustatten, dass es sich nicht um authentische Aufnahmen handelt, und sie in die Lage zu versetzen, selbstbestimmt die Informations- und Kommunikationsplattformen im digitalen Raum zu nutzen.[68] Wichtige Ausnahmen von der Kennzeichnungspflicht sind insbesondere zur Ausübung der Meinungs-, Kunst- und Forschungsfreiheit vorgesehen. Indessen ist die spezifi-

64 *Europäische Kommission*, Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final.

65 *Ebert/Spiecker gen. Döhmann*, NVwZ 2021, 1188 (1191).

66 *Ebert/Spiecker gen. Döhmann*, NVwZ 2021, 1188 (1191).

67 S. *Heesen* u. a., KI-Systeme und die individuelle Wahlentscheidung, 2021, 28; *Kalbhenn*, ZUM 2021, 663 (670).

68 Vgl. Begr. KIVO-E, S. 17.

sche Regelung hinsichtlich der Normadressaten zu hinterfragen. Statt einer nur für Nutzer geltenden Pflicht wäre für eine gesteigerte Wirksamkeit die Aufnahme von Systemgestaltern in den Adressatenkreis denkbar, sodass Deepfakes stets mit einem digitalen Wasserzeichen versehen werden müssten. Damit die im KIVO-E vorgeschlagenen Transparenzregelungen in der Praxis tatsächlich wirksam sind, muss es außerdem gelingen, eine effektive Kontrolle der Umsetzung und funktionierende Durchsetzungsstrukturen zu etablieren.

5 Erhöhte externe Kontrolle durch den Digital Services Act?

Auf EU-Ebene nimmt sich der Vorschlag der EU-Kommission für ein „Gesetz über digitale Dienste" (Digital Services Act)[69] der Aufgabe an, mit einem komplexen Regelwerk einen klaren Transparenz- und Rechenschaftsrahmen für Online-Plattformen zu schaffen und eine Rechtszersplitterung im digitalen Binnenmarkt zu verhindern. Der Entwurf weist große Ähnlichkeiten zu den nationalen Compliance- und Transparenzvorgaben im Netzwerkdurchsetzungsgesetz und Medienstaatsvertrag auf, insbesondere im Hinblick auf den Umgang mit illegalen Inhalten. Was nicht gesetzeswidrige Desinformation betrifft, sind keine spezifischen Pflichten bezüglich Maßnahmen wie Entfernung oder Downranking enthalten. Auch bleiben die Grundsätze der Host-Provider-Haftung aus der E-Commerce-Richtlinie, für illegale Inhalte nur zu haften, sofern sie in Kenntnis gesetzt werden, und im Übrigen keine von ihnen übermittelten oder gespeicherten Informationen aktiv überwachen zu müssen, im Kern unangetastet.[70]

5.1 Desinformation als systemisches Risiko

Indessen könnten die gemäß Art. 25 ff. DSA-E vorgesehenen Verpflichtungen für sehr große Online-Plattformen wie Facebook, den Missbrauch ihrer Systeme zu verhindern, indem sie systemische Risiken identifizieren, risikobasierte Maßnahmen durchführen und ihr Risikomanagementsystem

69 *Europäische Kommission*, Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates über einen Binnenmarkt für digitale Dienste (Gesetz über digitale Dienste) und zur Änderung der Richtlinie 2000/31/EG, COM(2020) 825 final.

70 Vgl. etwa ErwG 71 DSA-E.

von unabhängiger Seite prüfen lassen, zu einem nachhaltigeren, umfassenderen Lösungsansatz der Desinformationsbekämpfung beitragen. Die systemische Risikobewertung umfasst nach Art. 26 Abs. 1 DSA-E neben der Verbreitung illegaler Inhalte explizit vorsätzliche Manipulationen des Dienstes, darunter unauthentische und automatisierte Ausnutzungen des Dienstes, mit tatsächlichen oder absehbaren nachteiligen Auswirkungen auf Rechtsgüter wie die öffentliche Gesundheit oder die öffentliche Sicherheit. Adressiert sind damit ohne weiteres Bot-Netze und Fake-Accounts. Unklar ist jedoch, inwieweit Desinformationen in anderen Formen, die nicht Teil einer Kampagne sind, erfasst sind.[71] Da andere Desinformationen, die vielfach unkoordiniert aber mit erheblicher Reichweite auftreten, ebenfalls erhebliche Risiken für die gesellschaftliche Debatte und weitere in der Vorschrift genannten Rechtsgüter aufweisen, sprechen Sinn und Zweck der Vorschrift dafür, sie auch als systemische Risiken anzusehen.

Unter Berücksichtigung von Art. 26 Abs. 2 DSA-E müssen die Plattformen mithin insbesondere ihre Systeme der Content-Moderation und algorithmischen Empfehlungen hinsichtlich ihres Einflusses auf die Weiterverbreitung von Desinformation und sonstigen Manipulationen bewerten. Dass Desinformation auch auf kleineren Online-Plattformen ein ernstes Problem ist und hier ebenfalls Echokammern und Radikalisierung begünstigt werden können, adressiert der Entwurf jedoch nicht. Indessen ist positiv zu vermerken, dass der Katalog der Risiken insofern nicht einseitig formuliert ist, wie sie auch den beschriebenen, zu vollbringenden Balanceakt implizit aufgreifen, indem nachteilige Auswirkungen auf die Ausübung der Grundrechte, insbesondere die Meinungs- und Informationsfreiheit ausdrücklich als systemische Risiken benannt werden. Konkretisierungen der systemischen Risiken könnten weiteren Aufschluss geben, z. B. mit der Erwähnung des Risikos irrtümlicher oder ungerechtfertigter Sperrungen.[72]

Spiegelbildlich zu den systemischen Risiken werden als Risikominderungsmaßnahmen im nicht abschließenden Katalog des Art. 27 Abs. 1 lit. a DSA-E insbesondere die Anpassung der Content-Moderation und der Empfehlungssysteme benannt. Zudem sind von der Kommission als geeignete Maßnahmen zur Risikominderung von Desinformation Verhaltenskodizes und Krisenprotokolle vorgesehen (vgl. Art. 27 Abs. 1 i.V.m. Art. 35, 37 DSA-E).[73] Absehbar könnte der bereits existierende Verhaltenskodex zur Bekämpfung von Desinformation weiter gestärkt und ausdiffe-

71 *Dreyer* u. a., Desinformation, 2021, 37.
72 Vgl. *Europäische Kommission*, COM (2020) 825 final, deutsche Fassung, S. 14.
73 S. ErwG 68 DSA-E.

renziert werden, nicht zuletzt hinsichtlich Aufsichtsmöglichkeiten im Sinne einer Ko-Regulierung. Herausfordernd ist darüber hinaus, die Vorgaben zum Risikomanagement mit hinreichenden Rechenschaftspflichten und Durchsetzungskraft in der Praxis auszustatten. Insoweit ist zweifelhaft, ob die mindestens einmal jährlich durchzuführende Risikobewertung (Art. 26 Abs. 1 S. 1 DSA-E) und der umfassende, einmal jährlich in Zusammenarbeit mit der Kommission zu erbringende Bericht (Art. 27 Abs. 2 DSA-E) geeignet sind, um über das äußerst dynamische Geschehen auf den Plattformen im Zusammenhang mit Desinformation zu informieren und darauf zu reagieren. Vielversprechend ist die mindestens einmal jährliche Prüfung der Einhaltung der auferlegten Pflichten und Verpflichtungszusagen durch unabhängige Sachverständige gemäß Art. 28 DSA-E, aus der jeweils ein Prüfbericht mit etwaigen operativen Empfehlungen hervorgeht, über deren Umsetzung die Plattformen wiederum innerhalb eines Monats berichten müssen. Auch hier ist jedoch fraglich, ob ein einmal jährliches Audit ausreichend ist.

Jedenfalls sind mit den vorgeschlagenen Verpflichtungen wichtige Grundsteine für eine weiterreichende Kontrolle der sehr großen Online-Plattformen als auf Bundesebene de lege lata gelegt. Dies zeigt sich nicht zuletzt an dem gemäß Art. 31 Abs. 1 DSA-E vorgesehenen Datenzugang für den Koordinator für digitale Dienste, der sich auf die für die Überwachung und Bewertung der Einhaltung der Verordnung erforderlichen Daten und somit auf Geschäftsgeheimnisse wie die Prüfung von Algorithmen erstreckt.[74] Auf diese Weise sollen auch Daten zur Genauigkeit und Funktionsweise von algorithmischen Systemen der Content-Moderation eingesehen werden, die erheblich präziser sein dürften als die üblicherweise sehr vage und unbestimmt gehaltenen Informationen der Plattformen. Auch der für Forschende vorgesehene Datenzugang gemäß Art. 31 Abs. 2 DSA-E wird den Abbau der Wissensasymmetrien bis zu einem gewissen Grad vorantreiben.[75] Schließlich umfassen die Befugnisse der Kommission u. a. gemäß Art. 54 DSA-E Nachprüfungen vor Ort bei der betreffenden Online-Plattform und Erläuterungen zu Algorithmen sowie gemäß Art. 55 DSA-E den Erlass einstweiliger Maßnahmen unter den engen Voraussetzungen der Dringlichkeit und der Gefahr einer schwerwiegenden Schädigung der Nutzer.

74 S. ErwG 60 und 64 DSA-E.
75 S. ErwG 64 DSA-E.

5.2 Einschränkung und Offenlegung von Automatisierung

Den Einsatz automatisierter Mittel greift der DSA-E im Rahmen von Melde- und Abhilfeverfahren sowie der Begründung von Entscheidungen der Hosting-Anbieter und Online-Plattformen auf. So sind im Rahmen der „Notice-and-Action"-Mechanismen, bei denen Anbieter über durch Nutzende oder andere Akteure gemeldete Inhalte zu entscheiden haben, automatisierte Verfahren ohne menschliche Kontrolle nicht ausgeschlossen. Dies folgt im Umkehrschluss aus Art. 14 Abs. 6 Satz 2 DSA-E, der vorgibt, die Person, die einen Beitrag gemeldet hat, bei der Bestätigung über den Erhalt der Notifizierung auch über eine etwaige automatisierte Bearbeitung oder Entscheidungsfindung zu informieren. Entsprechende Informationspflichten über den Einsatz automatisierter Mittel sollen gemäß Art. 15 Abs. 1, Abs. 2 lit. c DSA-E außerdem bei der Begründung von Entscheidungen über entfernte oder gesperrte Inhalte gelten. Mithin schließen die Vorgaben vollautomatisierte Verfahren auch im Rahmen der Content-Moderation nicht aus. Allerdings wird der Einsatz insoweit voraussetzungsvoll, als Entscheidungen mit einer klaren und spezifischen Begründung spätestens im Zeitpunkt der Entfernung oder Zugangssperrung versehen sein müssen. Strengere Vorgaben sind hingegen bei Letztentscheidungen im Rahmen der internen Beschwerdeverfahren der Online-Plattformen wegen entfernter Inhalte und Konten vorgesehen. Hier ist mit Blick auf die skizzierten Risiken algorithmischer Entscheidungsfindung die explizite Vorgabe in Art. 17 Abs. 5 DSA-E, über Beschwerden von Nutzern gegen Entfernungen von Beiträgen und Nutzerkonten nicht ausschließlich automatisiert zu entscheiden, zu begrüßen.

Eine weitere erhebliche Transparenzsteigerung, auch gegenüber den Vorgaben des nationalen Rechtsrahmens, steht in Aussicht, soweit Hosting-Anbieter nach Art. 15 Abs. 4 DSA-verpflichtet werden sollen, Entscheidungen und Begründungen über Löschungen und Sperrungen von Inhalten in einer öffentlich zugänglichen Datenbank, die von der Kommission verwaltet wird, zu veröffentlichen. Erst die Zurverfügungstellung der anonymisierten Einzelfälle mit jeweiliger Begründung der Entscheidung ermöglicht es, die Frage des Over- und Underblocking zu prüfen und gesellschaftlich zu diskutieren.[76] Hinsichtlich der Lösch- und Sperrpraxis von Desinformation durch die Plattformen könnten diese Vorgaben ebenfalls für mehr Transparenz sorgen.

76 *Löber/Roßnagel*, MMR 2019, 71 (75); mit dieser Forderung bereits *Eifert*, NJW 2017, 1450 (1453).

6 Fazit und Ausblick

KI-basierte Detektions- und Filtersysteme in Social Networks sind unverzichtbare Werkzeuge zur Eindämmung von Desinformation. Staatliche Regulierungsoptionen sind im Hinblick auf Desinformation stark begrenzt, da die freiheitlich-demokratische Grundordnung bei nicht rechtswidrigen Äußerungen auf die Kraft der gesellschaftlichen Auseinandersetzung vertraut. Daher muss die Regulierung darauf gerichtet sein, für notwendige Rahmenbedingungen zu sorgen, zu denen insbesondere ein grundrechtsschonender und transparenter Umgang mit den KI-Systemen und der Content-Moderation mächtiger Online-Plattformen im digitalen Raum gehört. Für den rechtmäßigen Einsatz der KI-Systeme müssen die technischen Möglichkeiten und Grenzen hinreichend berücksichtigt und die widerstreitenden Grundrechtsbelange austariert werden. Neben einem verfahrensbasierten Grundrechtsschutz für Nutzende bleibt in vielen Fällen eine manuelle Nachkontrolle durch die Anbieter notwendig, um Overblocking nicht in Kauf zu nehmen.

Obwohl die Debatte über Desinformation in Social Networks seit den Aufdeckungen zu gezielten Falschmeldungen im US-Präsidentschaftswahlkampf 2016 intensiv geführt wird, bestehen beim Umgang mit dieser Problematik weiterhin erhebliche Informationsasymmetrien zwischen den Anbietern auf der einen und Online-Nutzenden, Politik, Forschung und Zivilgesellschaft auf der anderen Seite. Auch Unsicherheiten bezüglich der Reichweite der Befugnisse der Anbieter, in den öffentlichen Diskurs einzugreifen, sind nicht ausgeräumt. Auf Bundesebene wurden in jüngerer Vergangenheit erste zentrale Leitplanken zur gesetzlichen Einhegung der Befugnisse der Online-Plattformen und zur Transparenzsteigerung für Nutzende etabliert, wobei es den eingeführten Transparenzvorgaben teilweise noch an Schärfe und Überprüfbarkeit mangelt. Die jüngsten nationalen Gesetzesänderungen sowie die im Vergleich ambitionierteren und weltweit bislang einzigartigen Regulierungsbestrebungen auf Ebene der EU zeigen jedoch, dass die externe Kontrolle der Vorgänge auf den großen Online-Plattformen Fahrt aufnimmt. Als ein grundlegender Pfeiler ist die Unterscheidbarkeit von Mensch und Maschine im DSA-E und im KIVO-E fest verankert: sowohl, wenn es um automatisierte Entscheidungen über die Entfernung von desinformativen und anderen Inhalten geht, als auch bei der Offenlegung von potenziell manipulativ wirkenden Techniken wie Social Bots und Deepfakes. Da zu erwarten steht, dass die KI-basierte Erkennung von Desinformation und verwandten Manipulationstechniken künftig weitere Fortschritte erzielen wird, ist es umso bedeutsamer, Chan-

cen und Risiken dieser Techniken für die Gesellschaft und für die Kommunikationsgrundrechte rechtlich zu adressieren.

Literatur

Alba, Davey und Mac, Ryan (2021): Facebook, Fearing Public Outcry, Shelved Earlier Report on Popular Posts, *New York Times* vom 20.8.21, aktualisiert am 25.10.2021. URL: https://www.nytimes.com/2021/08/20/technology/facebook-popular-posts.html (besucht am 28.02.2022).

Cornils, Matthias (2019): Die Perspektive der Wissenschaft: AVMD-Richtlinie, der 22. Rundfunkänderungsstaatsvertrag und der „Medienstaatsvertrag" – Angemessene Instrumente für die Regulierungsherausforderungen? *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 2, S. 89-103.

Drexl, Josef (2017): Bedrohung der Meinungsvielfalt durch Algorithmen. *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 7, S. 529-543.

Dreyer, Stephan; Stanciu, Elena; Potthast, Keno Christopher und Schulz, Wolfgang (2021): Desinformation: Risiken, Regulierungslücken, und adäquate Gegenmaßnahmen. Wissenschaftliches Gutachten im Auftrag der Landesanstalt für Medien NRW. Düsseldorf: Landesanstalt für Medien NRW.

Dürig, Günter; Herzog, Roman und Scholz, Rupert (Hrsg.) (2021): *Grundgesetz Kommentar*, 95. Aufl., Stand Juli 2021. München: C.H.BECK.

Ebert, Andreas und Spiecker gen. Döhmann, Indra (2021): Der Kommissionsentwurf für eine KI-Verordnung der EU, Die EU als Trendsetter weltweiter KI-Regulierung. *Neue Zeitschrift für Verwaltungsrecht (NVwZ)*, Heft 16, S. 1188-1193.

Eifert, Martin (2017): Rechenschaftspflichten für soziale Netzwerke und Suchmaschinen. Zur Veränderung des Umgangs von Recht und Politik mit dem Internet. *Neue juristische Wochenschrift (NJW)*, Heft 20, S. 1450-1454.

Facebook Ireland Limited (Februar 2020): Stellungnahme zum Referentenentwurf eines Gesetzes zur Änderung des Netzwerkdurchsetzungsgesetzes, Einreichung für Facebook Ireland Limited. URL: https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/Stellungnahmen/2020/Downloads/021720_Stellungnahme_Facebook_RefE_NetzDG.pdf;jsessionid=93EE694999ACAFFD93DE4F972663D0A0.2_cid289?__blob=publicationFile&v=2 (besucht am 28.02.2022).

Feldwisch-Drentrup, Hinnerk und Kuhrt, Nicola (2019): Schlechte und gefährliche Gesundheitsinformationen. Wie sie erkannt und Patienten besser geschützt werden können. Gütersloh: Bertelsmann Stiftung. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/schlechte-und-gefaehrliche-gesundheitsinformationen/ (besucht am 28.02.2022).

Google (2022): YouTube-Community-Richtlinien und ihre Anwendung. Transparenzbericht. URL: https://transparencyreport.google.com/youtube-policy/removals (besucht am 28.02.2022).

Halvani, Oren; Heereman von Zuydtwyck, Wendy Freifrau; Herfert, Michael; Kreutzer, Michael; Liu, Huajian; Simo Fhom, Hervais-Clemence; Steinebach, Martin; Vogel, Inna; Wolf, Ruben; Yannikos, York; Zmudzinski, Sascha (2020): Automatisierte Erkennung von Desinformationen. In: Steinebach, Martin; Bader, Katarina; Rinsdorf, Lars; Krämer, Nicole und Roßnagel, Alexander (Hrsg.): *Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität*. Baden-Baden: Nomos, S. 101-148. DOI: doi.org/10.5771/9783748904816

Heesen, Jessica; Bieber, Christoph; Grunwald, Armin; Matzner, Tobias und Roßnagel, Alexander (2021): KI-Systeme und die individuelle Wahlentscheidung. Chancen und Herausforderungen für die Demokratie. Whitepaper. München: Lernende Systeme – Die Plattform für Künstliche Intelligenz. DOI: https://doi.org/10.48669/pls_2021-1

Heidtke, Aron (2020): *Meinungsbildung und Medienintermediäre, Vielfaltssichernde Regulierung zur Gewährleistung der Funktionsbedingungen freier Meinungsbildung im Zeitalter der Digitalisierung*. Baden-Baden: Nomos.

Hölig, Sascha; Hasebrink, Uwe und Behre, Julia (2021): Reuters Institute Digital News Report 2021 – Ergebnisse für Deutschland. Hamburg: Verlag Hans-Bredow-Institut, Juni 2021 (Arbeitspapiere des Hans-Bredow-Instituts, Projektergebnisse Nr. 58). DOI: https://doi.org/10.21241/ssoar.73637

Huesmann, Felix (2020): Qanon – der Aufstieg einer gefährlichen Verschwörungstheorie, RND vom 11.4.2020. URL: https://www.rnd.de/politik/qanon-der-aufstieg-einer-gefahrlichen-verschworungstheorie-ORTPE4D5YRFRZKVTMJBTFADJTY.html (besucht am 28.02.2022).

Ingold, Albert (2019). Governance of Algorithms. Kommunikationskontrolle durch „Content Curation" in sozialen Netzwerken. In: Unger, Sebastian und von Ungern-Sternberg, Antje (Hrsg.): *Demokratie und Künstliche Intelligenz*. Tübingen: Mohr Siebeck, S. 183-213.

Ipsen, Flemming; Zywietz, Bernd; Böndgen, Franziska; Hebeisen, Michael; Schneider, Sebastian; Schnellbacher, Jan und Wörner-Schappert, Michael (2021): Bericht Rechtsextremismus im Netz 2020/21, November 2021. Mainz: jugendschutz.net. URL: https://www.jugendschutz.net/fileadmin/daten/publikationen/lageberichte/bericht_2020_2021_rechtsextremismus_im_netz.pdf (besucht am 28.02.2022).

Isensee, Josef und Kirchhof, Paul (Hrsg.) (2005): *Handbuch des Staatsrechts, Band III, Demokratie – Bundesorgane*, 3. Aufl. Heidelberg: C.F. Müller.

Islam, Md Saiful; Sarkar, Tonmoy; Khan, Sazzad Hossein; Mostofa Kamal, Abu-Hena; Hasan, S M Murshid; Kabir, Alamgir; Yeasmin, Dalia; Islam, Mohammad Ariful; Amin Chowdhur, Kamal Ibne; Anwar, Kazi Selim; Chughtai, Abrar Amad; Seale, Holly (2020): COVID-19-Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *American Journal of Tropical Medicine and Hygiene (Am J Trop Med Hyg.)*, 2020 Oct;103(4):1621-1629. DOI: 10.4269/ajtmh.20-0812.

Jaster, Romy und Lanius, David (2020): Schlechte Nachrichten: „Fake News" in Politik und Öffentlichkeit. In: Hohlfeld, Ralf; Harnischmacher, Michael; Heinke, Elfi; Lehner, Lea und Sengl, Michael (Hrsg.): *Fake News und Desinformation: Herausforderungen für die vernetzte Gesellschaft und die empirische Forschung*. Baden-Baden: Nomos, S. 245-267. DOI: https://doi.org/10.5771/9783748901334

Kahl, Wolfgang; Waldhoff, Christian und Walter, Christian (Hrsg.) (2021): *Bonner Kommentar zum Grundgesetz*. Loseblattsammlung, Stand des Gesamtwerks: 214. Aktualisierung Dezember 2021. Heidelberg: C.F. Müller.

Kajewski, Marie-Christine (2017): Wahrheit und Demokratie in postfaktischen Zeiten. *Zeitschrift für Politik (ZfP)*, 64. Jg. 4/2017, S. 454-467.

Kalbhenn, Jan Christopher (2021): Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung. *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 8/9, S. 663-674.

Kastl, Graziana (2016): Filter – Fluch oder Segen? Möglichkeiten und Grenzen von Filtertechnologien zur Verhinderung von Rechtsverletzungen. *Gewerblicher Rechtsschutz und Urheberrecht (GRUR)*, Heft 7, S. 671-678.

Kreye, Andrian (2021): Warum die Rohingya Facebook verklagen, *Süddeutsche Zeitung* vom 7.12.2021, URL: https://www.sueddeutsche.de/politik/rohingya-facebook-meta-klage-1.5482494.

Kringiel, Danny (2021): Pandemie? Welche Pandemie? *Spiegel Online* vom 16.2.2021, URL: https://www.spiegel.de/geschichte/corona-aids-und-krebs-leugner-die-krankheit-einfach-wegglauben-a-c6b27102-3ad0-4ac0-aff7-0b96a4ae5db1.

Kühling, Jürgen (2021): „Fake News" und „Hate Speech" – Die Verantwortung der Medienintermediäre zwischen neuen NetzDG, MStV und Digital Services Act, *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 6, 461-472.

Ladeur, Karl-Heinz und Gostomzyk, Tobias (2018): Das Medienrecht und die Herausforderung der technologischen Hybridisierung. Eine Kommentierung der Regelungen zu Medienintermediären im Entwurf des Medienstaatsvertrags der Länder. *Kommunikation und Recht (K&R)*, Heft 11, S. 686-693.

Lazer, David M. J.; Baum, Matthew; Benkler, Yochai und Berinsky, Adam J. u. a. (2018): The Science of Fake News. *Science* 359(6380), S. 1094-1096. DOI: https://doi.org/10.1126/science.aao2998

Löber, Lena Isabell und Roßnagel, Alexander (2019): Das Netzwerkdurchsetzungsgesetz in der Umsetzung. *Zeitschrift für IT-Recht und Recht der Digitalisierung (MMR)*, Heft 2, S. 71-76.

Löber, Lena Isabell und Roßnagel, Alexander (2019): Kennzeichnung von Social Bots. Transparenzpflichten zum Schutz integrer Kommunikation. *Zeitschrift für IT-Recht und Recht der Digitalisierung (MMR)*, Heft 8, S. 493-498.

Löber, Lena Isabell und Roßnagel, Alexander (2020): Desinformation aus der Perspektive des Rechts. In: Steinebach, Martin; Bader, Katarina; Rinsdorf, Lars; Krämer, Nicole und Roßnagel, Alexander (Hrsg.): *Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität*. Baden-Baden: Nomos, S. 149-194. DOI: doi.org/10.5771/9783748904816

Löber, Lena Isabell (28.10.2021): KI-Lösungen gegen Desinformation in Social Networks – Fragen des grundrechtskonformen und transparenten Einsatzes. URL: https://forum-privatheit.de/blog/2021/10/28/ki-loesungen-gegen-desinformation-in-social-networks-fragen-des-grundrechtskonformen-und-transparenten-einsatzes/ (besucht am 28.02.2022).

Martini, Mario (2019): *Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz*. Berlin und Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-662-59010-2

Mengden, Martin (2018): *Zugangsfreiheit und Aufmerksamkeitsregulierung. Zur Reichweite des Gebots der Gewährleistung freier Meinungsbildung am Beispiel algorithmengestützter Zugangsdienste im Internet*. Tübingen: Mohr Siebeck.

Merten, Detlef und Papier, Hans-Jürgen (Hrsg.) (2011): *Handbuch der Grundrechte in Deutschland und Europa, Band IV: Grundrechte in Deutschland - Einzelgrundrechte I*. Heidelberg: C.F. Müller.

Mitsch, Lukas (2019): Soziale Netzwerke und der Paradigmenwechsel des öffentlichen Meinungsbildungsprozesses. *Deutsches Verwaltungsblatt (DVBl)*, Heft 13, S. 811-818.

Raue, Benjamin und Steinebach, Martin (2020): Uploadfilter – Funktionsweisen, Einsatzmöglichkeiten und Parametrisierung. *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 5, S. 355-364.

Reinhardt, Jörn und Yazicioglu, Melisa (2020): Grundrechtsbindung und Transparenzpflichten sozialer Netzwerke. *Tagungsband Herbstakademie 2020 (DSRITB)*, Heft 1, S. 819-833.

Roßnagel, Alexander (2020): Technik, Recht und Macht. Aufgabe des Freiheitsschutzes in Rechtsetzung und -anwendung im Technikrecht. *Zeitschrift für IT-Recht und Recht der Digitalisierung (MMR)*, Heft 4, S. 222-228.

Roßnagel, Alexander (29.10.2021): Datenspenden für KI – Vertrauen nur mit Grundrechtsvorsorge. URL: https://forum-privatheit.de/blog/2021/10/29/datenspenden-fuer-ki-vertrauen-nur-mit-grundrechtsvorsorge/ (besucht am 28.02.2022).

Schwartmann, Rolf; Hermann, Maximilian und Mühlenbeck, Robin (2020): *Transparenz bei Medienintermediären*. Herausgegeben von Medienanstalt Hamburg/Schleswig-Holstein. Leipzig: VISTAS.

Specht-Riemenschneider, Louisa (2019): Löschung beleidigender Äußerungen auf Facebook. Anmerkung zu EuGH, Urteil vom 3.10.2019 – C-18/18 – Glawischnig-Piesczek. *Zeitschrift für IT-Recht und Recht der Digitalisierung (MMR)*, Heft 12, S. 801-802.

Stark, Birgit; Stegmann, Daniel; mit Magin, Melanie und Jürgens, Pascal (2020): Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse. Berlin: AlgorithmWatch.

Tagesschau (2021): „Facebook stellt Profite über die Menschen“, *Tagesschau* vom 5.10.2021, URL: https://www.tagesschau.de/ausland/amerika/facebook-anhoerung-whistleblowerin-101.html.

Vosoughi, Soroush; Roy, Deb und Aral, Sinan (2018): The Spread of True and False News Online. *Science* 359(6380), S. 1146-1151. DOI: https://doi.org/10.1126/science.aap9559

Zimmer, Anja (2019): Smart Regulation: Welche Antworten gibt der Medienstaatsvertrag auf die Regulierungsherausforderungen des 21. Jahrhunderts? – Ein Blick aus der Regulierungspraxis. *Zeitschrift für Urheber- und Medienrecht (ZUM)*, Heft 2, S. 126-130.