

Stellungnahme gemäß § 27a BVerfGG von D64 – Zentrum für Digitalen Fortschritt

Aktenzeichen: 1 BvR 2152/25

Zu D64 – Zentrum für Digitalen Fortschritt

D64 ist ein gemeinnütziger und unabhängiger Verein. Unsere über 800 Mitglieder begreifen die digitale Transformation als Chance, das Miteinander unserer modernen Gesellschaft zu verbessern. Wir gestalten in 13 Arbeitsgruppen die gesellschaftliche, ökologische, technologische und politische Entwicklung konstruktiv, kritisch und kreativ mit.

Die Grundwerte Freiheit, Gerechtigkeit und Solidarität durch eine progressive Digitalpolitik zu verwirklichen, ist unser Ziel. Dafür wirken wir mithilfe der breitgefächerten Expertise unserer Mitglieder als unabhängiger Verein, der in allen Themenbereichen der Digitalisierung vordenkt und Impulse gibt.

Übersicht

1. Welche Bedeutung kommt Online-Plattformen (Art. 3 Buchst. i DSA) und auch sehr großen Online-Plattformen (Art. 33 Abs. 4 Unterabs. 1 DSA) für den Meinungsdiskurs und die Meinungsbildung in demokratischen und pluralistischen Gesellschaften zu?.....3
2. Tragen die von Online-Plattformen verwandten Vorgaben/Bedingungen über die Festlegung der Zulässigkeit von veröffentlichten Inhalten der Nutzer (allgemeine Geschäftsbedingungen, Nutzungsbedingungen, Verhaltensrichtlinien u. a.) zu einem der Meinungsäußerungsfreiheit gerecht werdenden Meinungsdiskurs bei?.....9
3. Wie und unter Einsatz welcher Methoden identifizieren Online-Plattformen insbesondere nach ihren Nutzungsbedingungen vertragswidrige und rechtswidrige Nutzerinhalte?.....13
4. Wie bewerten Sie das Vorgehen der Anbieter der Online-Plattformen gegen Desinformationen?.....22
5. Tragen die Maßnahmen der Inhalte-Moderation durch Online-Plattformen zur Gewährleistung des Rechts auf freie Meinungsäußerung angemessen bei?.....27

Kurzzusammenfassung

Sehr große Online-Plattformen wie X, Meta, TikTok oder LinkedIn sind keine neutralen Intermediäre, sondern zentrale Infrastrukturen der gesellschaftlichen Kommunikation. Durch algorithmische Kuratierung und Moderationsentscheidungen gestalten sie aktiv mit, welche Äußerungen öffentliche Reichweite erhalten. Diese systemische Relevanz begründet erhöhte rechtliche Anforderungen, etwa durch den Digital Services Act (DSA). Deren Durchsetzung in der Praxis bleibt jedoch erheblich hinter den Erwartungen zurück.

Mehr als 99 % aller Moderationsmaßnahmen auf VLOPs ergehen nicht wegen Rechtswidrigkeit, sondern wegen Verstößen gegen plattformeigene Nutzungsbedingungen. Der DSA verpflichtet Plattformen zwar, dabei Grundrechte zu berücksichtigen. Angesichts des hohen Automatisierungsgrads bei Moderationsentscheidungen ist jedoch zu bezweifeln, ob grundrechtskonforme Abwägungen tatsächlich getroffen werden. Die hohe Erfolgsquote von Nutzenden vor außergerichtlichen Streitbeilegungsstellen (ca. 52 %) zeigt, dass Moderationsentscheidungen der Meinungsfreiheit häufig nicht gerecht werden.

Auch bei Maßnahmen gegen Desinformation muss das Vorgehen verhältnismäßig sein. Dabei ist zwischen rechtswidrigen und legalen Inhalten zu unterscheiden. Erstere müssen entfernt werden, bei Letzteren sind mildere Maßnahmen wie Downranking, Demonetarisierung oder Faktenchecks grundsätzlich vorzuziehen. Besondere Zurückhaltung ist bei Inhalten geboten, die sich machtkritisch mit politischen Institutionen auseinandersetzen und keine unmittelbare Gefahr für die Gesundheit Dritter darstellen.

Die Inhaltsmoderation der (sehr großen) Online-Plattformen trägt dem Recht auf freie Meinungsäußerung derzeit nicht ausreichend Rechnung. Intransparente Algorithmen, kaum nachweisbare Shadowbanning-Praktiken und unzureichender Forschungsdatenzugang verhindern eine kritische Überprüfung der Machtakkumulation der globalen, gewinnorientierten Social-Media-Plattformen.

1. Welche Bedeutung kommt Online-Plattformen (Art. 3 Buchst. i DSA) und auch sehr großen Online-Plattformen (Art. 33 Abs. 4 Unterabs. 1 DSA) für den Meinungsdiskurs und die Meinungsbildung in demokratischen und pluralistischen Gesellschaften zu?

Online-Plattformen und insbesondere sehr große Online-Plattformen (nachfolgend VLOPs) haben sich von technischen Intermediären zu einer zentralen Infrastruktur in der gesellschaftlichen Kommunikation entwickelt. Sie agieren hierbei nicht allein als neutrale Darsteller öffentlicher Meinungen, sondern konstituieren und kuratieren diese durch algorithmische Sortierung und der Durchführung von Inhaltmoderation. Sie profitieren von Netzwerk- und Lock-In-Effekten, die ihre marktbeherrschende Stellung verfestigen und Nutzenden den Wechsel zu alternativen und kleineren Plattformen erschweren.

Angesichts einer umfassenden Verlagerung des Medienkonsums hin zu digitalen Räumen sind sie für die individuelle Meinungsbildung und den demokratischen Diskurs zunehmend systemrelevant geworden.¹ Die individuelle Möglichkeit, Meinungen zu äußern, sowie der Zugang zu vielfältigen Informationen hängen faktisch in hohem Maße von der Gestaltung dieser Plattformen und ihren Funktionen sowie ihren Regeln und Moderationsentscheidungen ab.

Diese Entwicklung spiegelt sich auch in der aktuellen EU-Gesetzgebung wider. Aufgrund dieser systemischen Relevanz hat der Digital Services Act (DSA) umfassende Verpflichtungen für VLOPs sowie sehr große Suchmaschinen eingeführt. Darüber hinaus ist der DSA Teil eines umfassenden Rechtsrahmens für die Informationsgesellschaft: So adressiert die Datenschutzgrundverordnung die Verpflichtungen bei der Verarbeitung personenbezogener Daten, das Gesetz über Digitale Märkte (DMA) die ökonomisch-

¹ Die Medienanstalten (2024) Medienvielfaltsmonitor, abgerufen am 19.02.2026: https://www.lfk.de/fileadmin/PDFs/Publikationen/Studien/MedienVielfaltsMonitor/Medienvielfaltsmonitor_2024.pdf, S. 28; Die Medienanstalten (2023) Intermediäre und Meinungsbildung, abgerufen am 19.02.2026: https://www.die-medienanstalten.de/fileadmin/user_upload/die_medienanstalten/Forschung/Intermediaere_und_Meinungsbildung/Intermedi%C3%A4re_Meinungsbildung_2023-II.pdf.

marktbeherrschende Stellung der Torwächter, die auch einige der als sehr große Online-Plattformen designierte Dienste umfasst, und die KI-Verordnung Risiken, die von KI-Systemen ausgehen können. Der DSA stellt einen umfassenden Rechtsrahmen für Vermittlungsdienste auf. Dieser beinhaltet neben Regelungen zur Haftung von Anbietern auch umfangreiche Sorgfaltspflichten für ein sichereres und transparenteres Online-Umfeld sowie die Verpflichtung zum Umgang mit systemischen Risiken. Der Umgang mit bestimmten Kategorien von Inhalten wird ferner durch weitere europäische Rechtsakte wie die Terroristische-Online-Inhalte-Verordnung (TCO-VO), die Verordnung zu Transparenz und Targeting von politischer Werbung (TTPW-VO), die Urheberrechtsrichtlinie (DSM-RL) sowie die geplante Verordnung zur Prävention und Bekämpfung des sexuellen Missbrauchs von Kindern (CSAM-VO) vorgegeben.

Funktionale Differenzierung der Plattformen, insbesondere der VLOPs

Der Status als VLOP knüpft nach Art. 33 DSA an eine quantitative Schwelle von 45 Mio. aktiven Nutzenden an.² Diese rein quantitative Betrachtung bildet jedoch die qualitative Bedeutung für den öffentlichen Diskurs nur unzureichend ab. So können Plattformen für spezifische gesellschaftliche Teilbereiche eine hohe Relevanz entfalten, die sich nicht alleine aus Nutzendenzahlen ablesen lässt.

Für eine angemessene Bewertung der Bedeutung der Plattformen ist daher eine funktionale Differenzierung erforderlich:

Plattformen des öffentlichen Diskurses

Zu dieser Kategorie zählen Dienste wie X (vormals Twitter), Angebote von Meta (Facebook, Instagram), YouTube, TikTok, aber auch Snapchat oder LinkedIn. Diese Plattformen fungieren als digitale „Marketplace of Ideas“ und bieten abseits anderer Diskursangebote, etwa Nachrichtenportalen, niedrighschwellige Beteiligungsmöglichkeiten für politische und gesellschaftliche Debatten. Sie stellen regelmäßig den Kern von Debatte um die Bedeutung von „Social Media“-Plattformen dar und sind zentrale Treiber eines Wandels der öffentlichen Kommunikation von one-to-many zu many-to-many.³

² Eine vollständige Liste der VLOPs kann unter <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses> eingesehen werden.

³ Sabine Fischer (2025): Geomax 31: Demokratie und Social Media – über die Wirkung von sozialen Netzwerken, <https://www.max-wissen.de/max-hefte/geomax-31-demokratie-und-social-media/>.

Empirische Erhebungen zeigen die Bedeutung dieser Plattformen für die öffentliche Meinungsbildung. Der Reuters Institute Digital News Report von 2024 zeigt, dass in Deutschland 15 % der Befragten Nachrichten hauptsächlich aus sozialen Medien beziehen; bei den 18-24-Jährigen sind es 35 %. In dieser Altersgruppe nutzen 16 % der Befragten soziale Medien sogar als einzige Quelle für Nachrichten.⁴

Es gibt unterschiedliche Nutzungsformen der verschiedenen Dienste, die auch innerhalb einer einzelnen Plattform zusammenfallen können. Die folgenden Kategorien sind dabei lediglich als Heuristik zu verstehen und können keine trennscharfe, eindeutige Abgrenzung leisten:

1. „Klassische“ soziale Netzwerke, wie Facebook, Instagram und Snapchat, werden primär für Kommunikation in einem bekannten sozialen Umfeld, etwa mit Freund:innen, Familie, Bekannten oder persönlichen Alltagskontakte genutzt beziehungsweise sind zumindest ursprünglich hierfür konzipiert;
2. Öffentliche Diskursplattformen, wie X, TikTok, YouTube sowie in Teilen (bzw. auf einzelne Funktionen bezogen) Instagram oder Snapchat, richten sich an eine breite, eher unbekannte Öffentlichkeit. Die Sichtbarkeit hängt hier oft weniger von persönlichen Beziehungen als vielmehr von Follower-Zahlen und algorithmischer Kuratierung ab;
3. Berufliche Netzwerke, wie LinkedIn oder Xing, sind primär auf professionelle Kontexte ausgerichtet. Ziel der Sichtbarkeit auf diesen Plattformen ist es, potenzielle Arbeitgeber:innen, Geschäftspartner:innen oder Kund:innen anzusprechen, persönliche Expertise sichtbar zu machen und berufliche Netzwerke aufzubauen oder zu festigen. Die grundsätzliche Relevanz beruflicher Online-Netzwerke ist dabei stark von der jeweiligen Branche abhängig.

⁴ Behre, Julia; Hölig, Sascha; Judith Möller (2024): Reuters Institute Digital News Report 2024 – Ergebnisse für Deutschland. Hamburg: Verlag Hans-Bredow-Institut, Juni 2024 (Arbeitspapiere des Hans-Bredow-Instituts | Projektergebnisse Nr. 72), S.5 <https://doi.org/10.21241/ssoar.94461>

Informationsplattformen: Sonderfall Wikipedia

Wikipedia nimmt unter den Plattformen eine Sonderrolle ein. Als öffentliche und frei verfügbare Enzyklopädie ohne kommerzielle Orientierung dient sie als wichtige Faktenbasis für weiterführende Diskurse. Aufgrund der freien Lizenz der dort veröffentlichten Inhalte stellt sie ferner eine der wichtigsten Datengrundlagen für Zusammenfassungen für eine Vielzahl von Suchmaschinen dar.

Wikipedia unterscheidet sich grundlegend von anderen als VLOP eingestuften Anbietern. Entscheidungen über die Moderation von Inhalten erfolgen gemeinschaftlich, transparent und ohne wirtschaftliche Interessen. Anders als bei kommerziellen Plattformen erfolgt keine Optimierung auf typische Parameter wie Aufrufzahlen, Interaktion oder Aufmerksamkeitsbindung zugunsten Werbetreibender.

Online-Marktplätze, Bewertungsportale und Verzeichnisse

Auch Online-Marktplätze oder Bewertungsportale, wie Amazon, Zalando, Booking.com oder Google Maps, die primär auf den wirtschaftlichen Austausch von Waren und Dienstleistungen ausgerichtet sind, beziehungsweise diesen durch Bewertungen aktiv mitgestalten, sind für die öffentliche Meinungsbildung relevant. So erfolgen Konsumententscheidungen häufig aufgrund von Rezensionen und sind damit Teil sozialer Meinungsbildung. Zugleich beeinflussen Entscheidungen über die Sichtbarkeit von Produkten oder anderen Angeboten im Online-Handel, etwa Literatur oder Kulturangebote, die Informationsgrundlagen von Nutzenden und damit auch die daraus folgende Handlungsfreiheit, auch jenseits unmittelbarer politischer Diskurse.

Die Relevanz dieser Anbieter ist je nach Moderationsansatz unterschiedlich zu bewerten. So können etwa im Kontext von Rezensionen negative Anreize für die Meinungsäußerung entstehen: Bots oder bezahlte Dienstleister überschwemmen Plattformen mit positiven Bewertungen, um Gewerbetreibenden ein höheres Ranking zu verschaffen. Auch das Löschen negativer, aber legitimer Rezensionen verzerrt das Angebot. Obwohl sich viele der Anbieter als reine Vermittlungsplattformen darstellen, sollte ihre Bedeutung für die öffentliche Meinungsbildung daher nicht unterschätzt werden.

Risiken für die öffentliche Meinungsbildung

Die zentrale Bedeutung von Online-Plattformen für den öffentlichen Diskurs geht mit spezifischen Risiken für die freie Meinungsbildung einher. Die Auswahl der ausgespielten Inhalte beruht auf kommerziellen Interessen der Plattformbetreiber, die in der Regel nicht den Bedürfnissen eines demokratischen Diskurses entsprechen. Ziel ist nicht Information, Unterhaltung oder Nachrichtenwert, sondern eine möglichst umfassende Bindung der Aufmerksamkeit sowie Interaktion mit der Plattform (engl. „Engagement“), was regelmäßig besonders emotionale und aufregende Inhalte bevorteilt.⁵ Eine solche Kuratierungs-Logik fördert die Verbreitung von Desinformation, Hassrede sowie eine toxische Streitkultur.

Umgang mit (Fehl-)Informationen

Die in der öffentlichen Debatte häufig herangezogene These abgeschlossener „Filterblasen“ oder „Echokammern“ wird in der Wissenschaft inzwischen differenzierter betrachtet.⁶ Empirische Studien legen nahe, dass Nutzende auch auf Plattformen regelmäßig mit unterschiedlichen Themen und Perspektiven in Berührung kommen. Dennoch entscheiden algorithmische Kuratierung und Plattformdesign maßgeblich darüber, wie die Begegnung mit gegensätzlichen Positionen erfolgt.

Darüber hinaus haben laut Digital News Report 42 % der Erwachsenen Internetnutzenden Bedenken, ob sie Falschmeldungen von Fakten unterscheiden können. 26 % der Befragten geben an, bereits mit falschen oder irreführenden Informationen zum Thema Migration oder Politik in Kontakt gekommen zu sein.⁷ Solche Wahrnehmungen beeinflussen das Vertrauen in öffentliche Kommunikation.

⁵ Smitha Milli, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, Anca D Dragan: Engagement, user satisfaction, and the amplification of divisive content on social media, PNAS Nexus, März 2025, <https://doi.org/10.1093/pnasnexus/pgaf062>.

⁶ Bruns, Axel (2019): Filter bubble. *Internet Policy Review*, 8(4), <https://doi.org/10.14763/2019.4.1426>.

⁷ Behre, Julia; Hölig, Sascha; Judith Möller (2024): Reuters Institute Digital News Report 2024 – Ergebnisse für Deutschland. Arbeitspapiere des Hans-Bredow-Instituts | Projektergebnisse Nr. 72, Juni 2024, S.6 <https://doi.org/10.21241/ssoar.94461>.

Asymmetrie in der Beteiligung auf Diskursplattformen

Ein oft unterschätzter Faktor sind unterschiedliche Partizipationsdynamiken: Nur ein vergleichsweise kleiner Teil der Nutzenden erstellt oder kommentiert Inhalte, während die große Mehrheit überwiegend passiv konsumiert.⁸ Moderationsmaßnahmen, die sich gegen Inhalteersteller richten, haben somit auch Effekte über diese Gruppe hinaus, weil sie zugleich den Informationszugang vieler passiver Nutzender beeinflussen und den Rahmen öffentlich sichtbarer Diskussionen verändern. Der Schutz der aktiven Meinungsäußerung auf VLOPs ist daher auch mit der Sicherung des passiven Informationszugangs verbunden.

Fehlender Forschungsdatenzugang

Der DSA zielt mit verbessertem Datenzugang (Art. 40) für Wissenschaftler:innen darauf ab, die Erforschung solcher Dynamiken und deren Auswirkungen auf die öffentliche Meinungsbildung und Demokratie bei VLOPs zu erleichtern. In der Praxis erschweren die Plattformen Wissenschaftler:innen jedoch häufig den Zugang zu qualitativ hochwertigen Daten, sodass gerichtliche Verfahren zur Durchsetzung des Anspruchs der Wissenschaftler:innen erforderlich werden.⁹ Damit bleiben zentrale Voraussetzungen für eine unabhängige Überprüfung der Bedeutung von Plattformen auf öffentliche Diskurse unerfüllt.

⁸ Oswald, Lisa; Schulz, Will; Hertwig, Ralph; Lazer, David & Stier, Sebastian (2025): The Tip of the Iceberg: How the Social Media Production-Consumption Gap Distorts Public Opinion for Citizens and Researchers [Preprint], SocArXiv, https://doi.org/10.31235/osf.io/frcv5_v1.

⁹ European Commission (2025): Commission preliminarily finds TikTok and Meta in breach of their transparency obligations under the Digital Services Act, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2503; Gesellschaft für Freiheitsrechte (2026): Erfolg für GFF und DRI vor Berliner Kammergericht: Social-Media-Plattform X muss Daten für Forschung herausgeben, <https://freiheitsrechte.org/ueber-die-gff/presse/pressemitteilungen-der-gesellschaft-fur-freiheitsrechte/erfolg-fuer-gff-und-dri-vor-berliner-kammergericht-social-media-plattform-x-muss-daten-fuer-forschung-herausgeben>.

2. Tragen die von Online-Plattformen verwandten Vorgaben/Bedingungen über die Festlegung der Zulässigkeit von veröffentlichten Inhalten der Nutzer (allgemeine Geschäftsbedingungen, Nutzungsbedingungen, Verhaltensrichtlinien u. a.) zu einem der Meinungsäußerungsfreiheit gerecht werdenden Meinungsdiskurs bei?

Die Regulierung von Inhalten auf Online-Plattformen bewegt sich stets im Spannungsverhältnis zwischen Over- und Underblocking. Overblocking (übermäßige Moderation von Inhalten) birgt Gefahren für die Kommunikationsfreiheiten, insbesondere wenn Menschen bereits aus der Befürchtung einer Moderationsmaßnahme rechtmäßige Äußerungen unterlassen (*chilling effects*). Doch auch Underblocking (zu geringe Moderationsmaßnahmen) kann die Kommunikationsfreiheiten beeinträchtigen. Eine sehr rohe Debattenkultur führt nachweislich dazu, dass sich Menschen aus dem Diskurs zurückziehen (*silencing effects*).¹⁰ Zudem hat die Verbreitung von Desinformation erhebliche Auswirkungen auf den gesellschaftlichen Diskurs, etwa durch Polarisierung.¹¹

Weder zu starkes Overblocking noch erhebliches Underblocking sind wünschenswert – beiden Gefahren muss mit kontextsensibler und differenzierter Moderation Rechnung getragen werden. Mit diesem Spannungsverhältnis muss zum einen der Gesetzgeber

¹⁰ Das NETTZ, Gesellschaft für Medienpädagogik und Kommunikationskultur, HateAid und Neue deutsche Medienmacher*innen als Teil des Kompetenznetzwerks gegen Hass im Netz (Hrsg.) (2024): Lauter Hass – leiser Rückzug. Wie Hass im Netz den demokratischen Diskurs bedroht – Ergebnisse einer repräsentativen Befragung, Berlin, https://kompetenznetzwerk-hass-imnetz.de/download_lauterhass.php; Markard, Nora, Bredler, Eva Maria (2021): Jeder schweigt für sich allein. Verfassungsblog, <https://verfassungsblog.de/alleine-schweigen>.

¹¹ Vasist, Pramukh Nanjundaswamy, Chatterjee, Debashis, & Krishnan, Satish (2024): The polarizing impact of political disinformation and hate speech: A cross-country configural narrative, Information Systems Frontiers 663–688, <https://doi.org/10.1007/s10796-023-10390-w>.

umgehen, wenn er Online-Plattformen reguliert sowie bestimmte Äußerungen untersagt. Doch auch Online-Plattformen selbst müssen es bei der Gestaltung sowie der skalierten Durchsetzung ihrer Nutzungsbedingungen in Moderationsmaßnahmen beachten.

Die Bedeutung der Nutzungsbedingungen ist erheblich: Mehr als 99 Prozent der Moderationsmaßnahmen sehr großer Online-Plattformen ergehen aufgrund Verstoßes gegen die Nutzungsbedingungen der Plattformen.¹² Dabei ist zu beachten, dass die Nutzungsbedingungen großer Online-Plattformen in aller Regel auch die Verbreitung rechtswidriger Inhalte verbieten, sodass ein illegaler Inhalt zumeist auch ein Verstoß gegen die Nutzungsbedingungen darstellt (siehe dazu auch Frage 3). Der hohe Prozentsatz an Moderationsentscheidungen, die auf den Nutzungsbedingungen von Plattformen basieren, lässt sich deshalb auch darauf zurückführen, dass die beiden Kategorien nicht exklusiv sind. Liegt ausnahmsweise kein Verstoß gegen die Regeln der Plattform trotz der Illegalität eines Inhalts nach staatlichem Recht vor, wird der Inhalt in der Regel nur lokal blockiert („Geoblocking“).

Gestaltung der Nutzungsbedingungen

Alle sehr großen Online-Plattformen etablieren vertraglich Nutzungsbedingungen, die festlegen, welche Inhalte auf ihren Diensten zulässig sind. Dabei gehen sie über das gesetzlich Verbotene hinaus und untersagen auch legale Inhalte, wie etwa die Darstellung von Nacktheit. Dies kann unterschiedliche Motivationen haben, oft richtet sich dies nach den antizipierten Erwartungen der Werbepartner als maßgebliche Kunden. Doch auch die Erwartungen der Nutzenden, deren Aufmerksamkeit Teil der Werbefinanzierung ist, spielen eine Rolle. Um die Erfahrung auf den Plattformen für Nutzende möglichst angenehm zu gestalten, werden auch Äußerungen untersagt, die sich außerhalb des strafrechtlich Untersagten bewegen, etwa Hassrede, die nicht die Schwelle zur strafbaren Beleidigung oder Volksverhetzung überschreitet. Weil die Plattformen nicht darlegen müssen, ob ein Inhalt gegen staatliches Recht oder lediglich gegen ihre Nutzungsbedingungen verstößt, können sie eigene Maßstäbe für ihre Moderationspraxis entwickeln und gewinnen so an Rechtssicherheit.

¹² EU Commission (2026), Two years of Digital Services Act allows 50 million content moderation decisions by platforms to be reversed, <https://digital-strategy.ec.europa.eu/en/news/two-years-digital-services-act-allows-50-million-content-moderation-decisions-platforms-be-reversed>.

Nutzungsbedingungen spielen ferner eine zentrale Rolle bei der Bekämpfung von Desinformation, die nur in bestimmten Fällen rechtswidrig ist (siehe hierzu Frage 4). Dies wird rechtlich auch von Art. 34, 35 DSA vorausgesetzt, der von VLOPs verlangt, bei der Gestaltung ihrer Nutzungsbedingungen negativen Auswirkungen auf die gesellschaftliche Debatte, beispielsweise durch Desinformation (ErwG 66 DSA), zu begegnen.

Zugleich ist zu beachten, dass diese Nutzungsbedingungen gestaltenden privaten Akteure kommerzielle Ziele verfolgen und keinem gesamtgesellschaftlichen Nutzen verpflichtet sind. Als private Betreiber der Plattformen steht es ihnen grundsätzlich frei, darüber zu entscheiden, welche Inhalte auf der von ihnen betriebenen Plattform zulässig sind. Immer häufiger zeigt sich zudem, dass die Gestaltung der Plattformen gezielt zu politischen Zwecken eingesetzt wird. So veränderte Meta im Januar 2025 die Regelungen zu Hassrede und beendete die Zusammenarbeit mit Fact-Checkern in den USA. Als Begründung nannte Mark Zuckerberg „institutionalisierte Zensur“, ein Schritt, der als Anpassung an die politischen Positionen von US-Präsident Donald Trump zu werten ist.¹³

Rechtliche Bindung an Grundrechte

Früher räumten die Nutzungsbedingungen den Plattformen häufig sehr weiten Entscheidungsspielraum bei der Moderation von Inhalten ein, dieser wurde jedoch zunächst durch die Rechtsprechung und schließlich durch die Vorgaben des DSA, insbesondere Art. 14 Abs. 4, eingeehrt. Demnach müssen Plattformen, egal welcher Größe, bei der Anwendung und Durchsetzung ihrer AGB Grundrechte berücksichtigen. Damit wird der Bedeutung von Plattformen, insbesondere VLOPs, für den gesellschaftlichen Diskurs und die politische Willensbildung Rechnung getragen (siehe Frage 1). Ihre AGB müssen transparent, nachvollziehbar und maschinenlesbar formuliert sein.

Je größer und bedeutender die Plattform ist, desto stärker wird die Pflicht, die in der EU-Charta garantierten Grundrechte wie das Recht auf freie Meinungsäußerung, die Freiheit und den Pluralismus der Medien sowie den Schutz vor Diskriminierung bei der Gestaltung (bei VLOPs) und Anwendung (gilt für alle Plattformen) der Nutzungsbedingungen zu

¹³ Deutschlandfunk (2025): Vor Trump eingeknickt?, <https://www.deutschlandfunk.de/meta-instagram-facebook-zuckerberg-faktencheck-beschaerungen-100.html>.

berücksichtigen. Art. 14 Abs. 4 DSA verlangt zudem ein sorgfältiges, objektives und verhältnismäßiges Vorgehen, was sicherstellen soll, dass alle Nutzenden vor Willkür geschützt werden.

Kleineren Plattformen, die beispielsweise nur auf bestimmte Communities ausgerichtet sind und *safe spaces* darstellen wollen, sind dagegen freier in der Gestaltung und Anwendung ihrer Nutzungsbedingungen. Dieser abgestufte Ansatz liegt auf einer Linie mit der Stadionverbots-Rechtsprechung des Bundesverfassungsgerichts¹⁴.

Praktische Umsetzung und Auswirkungen

Von der Ausgestaltung der Nutzungsbedingungen ist deren Anwendung zu trennen, bei der es nach wie vor erhebliche Schwächen gibt (siehe dazu insbesondere Frage 5). So entfernt beispielsweise TikTok sehr aktiv Inhalte bei Verstoß gegen die Nutzungsbedingungen. X ist, soweit Zahlen verfügbar sind, weitaus weniger aktiv und setzt eher Mechanismen wie Community Notes (Kontextualisierung durch andere Nutzenden) statt Entfernung ein (siehe Übersichten zu Frage 3). Wikipedia evaluiert streitige Inhalte durch interne Prozesse in der Community.

Die Frage, ob die Nutzungsbedingungen der Online-Plattformen zu einem der Meinungsäußerungsfreiheit gerecht werdenden Meinungsdiskurs beitragen, ist nicht eindeutig mit Ja oder Nein zu beantworten. Es ist jedoch wichtig zu betonen, dass sie für Online-Plattformen eine andere Funktion haben – sie sollen dazu beitragen, dass sich die für ihr Geschäftsmodell relevanten Akteure auf der Plattform wohlfühlen und dort Zeit verbringen. Außerdem sind sie rechtlich erforderlich.

Faktisch haben die privat gestalteten Nutzungsbedingungen weitreichende Auswirkungen darauf, welche Inhalte auf den Plattformen zulässig sind und welche nicht. Es ist daher richtig und geboten, dass sowohl die Gestaltung als auch die Durchsetzung (dazu näher Frage 5) der AGB kritisch an den Grundrechten der beteiligten Nutzenden gemessen wird. Diese Prüfung sollte umso strenger werden, je bedeutender die Plattform für den gesellschaftlichen Diskurs ist. Sie sollte dazu beitragen, ein ausgewogenes Maß an Moderation zu treffen, um sowohl Over- als auch Underblocking zu vermeiden.

¹⁴ BVerfG, 11. April 2018 – 1 BvR 3080/09.

3. Wie und unter Einsatz welcher Methoden identifizieren Online-Plattformen insbesondere nach ihren Nutzungsbedingungen vertragswidrige und rechtswidrige Nutzerinhalte?

Online-Plattformen nutzen unterschiedliche Methoden, Inhalte zu identifizieren und moderieren, die gegen Nutzungsbedingungen und/oder staatliches Recht verstoßen. Dabei kann sowohl nach dem Zeitpunkt der Moderationsmaßnahme als auch nach dem Grad der Automatisierung unterschieden werden:

Automatisierte und menschliche Moderation

Primär kommen automatisierte Verfahren zum Einsatz, um die Vereinbarkeit von Inhalten mit den Nutzungsbedingungen festzustellen.¹⁵ Je nach Art des Inhalts kommen hierfür KI-gestützte Text-, Audio- und Bilderkennung, Hashing-Verfahren für die Suche nach bekanntem Material oder auch wortbasierte Filter zum Einsatz. Obwohl Nutzer:innen von einer gezielten Abwertung ihrer Inhalte durch bestimmte Schlagworte berichten, eine umfassende Vermeidung bestimmter Begriffe auf Online-Plattformen zu beobachten ist,¹⁶ und auch Filterlisten bekannt geworden sind,¹⁷ leugnen viele Plattformen ihre Existenz.¹⁸

¹⁵ Trujillo, Amaury; Fagni, Tiziano. & Cresci, Stefano (2025), The DSA Transparency Database: Auditing Self-reported Moderation Actions by Social Media, <https://arxiv.org/abs/2312.10269>; Drolsbach, Chiara Patricia & Pröllochs, Nicolas (2024): Content Moderation on Social Media in the EU: Insights From the DSA Transparency Database, S. 6, <https://doi.org/10.1145/3589335.3651482>.

¹⁶ Steen, Ella, Yurechko, Kathryn, & Klug, Daniel (2023) You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231194586>.

¹⁷ Fehrensen, Martin; Berlin, Simon (2022) TikTok zensiert in Deutschland mit geheimen Filterlisten, *Social Media Watchblog*, <https://www.socialmediawatchblog.de/tiktok-zensiert-in-deutschland-mit-geheimen-filterlisten-instagram-fuehrt-einen-chronologischen-feed-ein-apple-arbeitet-an-abo-modell-fuer-hardware>.

¹⁸ Germain, Thomas (2025), The words you can't say on the internet, <https://www.bbc.com/future/article/20251118-the-words-you-cant-say-on-the-internet>.

Die Markierung eines Beitrags als rechts- oder nutzungsbedingungswidrig muss nicht zwingend zu einer klar erkennbaren Moderationsmaßnahmen wie der Löschung des Inhalts führen. Vielfach wird auch beobachtet oder vermutet, dass es zu einer verdeckten Verringerung der Reichweite eines Posts oder Accounts kommt (im Folgenden „Shadowbanning“). Diese Fälle sind für Nutzende schwer nachweisbar. Dabei ist jedoch festzuhalten, dass auch die algorithmische Herabstufung von Inhalten eine Moderationsmaßnahme im Sinne des DSA darstellt und Beschwerderechte sowie Informationspflichten auslöst (Art. 3 f) DSA). Dass entsprechende Daten derzeit nicht öffentlich verfügbar und auch im Rahmen bestehenden Rechts (DSA-Forschungsdatenzugang) von den Plattformen nur zurückhaltend und unter erheblichem Druck herausgegeben werden, trägt in der Kombination mit ihrem Einfluss auf den Informationskonsum der Nutzenden zum besonderen Risiko für die demokratische Meinungsbildung insgesamt bei.

Darüber hinaus erfolgt Mustererkennung nicht nur inhaltebasiert, sondern analysiert auch auffälliges („inauthentic“) Verhalten. Häufiges Posten über die gesamten 24 Stunden eines Tages, viele Inhalte gleicher Art wie andere Nutzerkonten und weitere Indikatoren können dazu führen, dass ein Konto gesperrt wird, weil vermutet wird, dass es sich um einen gegen die Nutzungsbedingungen verstoßenden Bot handelt.¹⁹

Zudem werden Moderationsmaßnahmen auch durch menschliche Moderator:innen durchgeführt, die oft unter prekären Arbeitsbedingungen,²⁰ ohne ausreichende Schulungen und psychologische Betreuung gewaltvolle Inhalte sichten müssen.²¹

¹⁹ Facebook (2020): Inauthentic Behavior Report, <https://about.fb.com/wp-content/uploads/2020/10/Inauthentic-Behavior-Report-October-2020.pdf>

²⁰ Gillespie, Tarleton (2018): Custodians Of The Internet: Platforms, Content Moderation, And The Hidden Decisions That Shape Social Media, S. 122; Block, Hans; Riesewieck, Moritz (2018): The Cleaners, <https://www.bpb.de/mediathek/video/273199/the-cleaners>.

²¹ Roberts, Sarah T. (2017): Social Media’s Silent Filter. *The Atlantic*, <https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796>.

Zeitpunkt der Moderation

Der Zeitpunkt der Moderation richtet sich regelmäßig auch nach der Art der Moderation. So erfolgen menschliche Moderationsentscheidungen bei VLOPs stets nach Veröffentlichung eines Beitrags, regelmäßig nach Meldung durch die Nutzenden. Anders ist es bei kleineren Plattformen, bei denen es durchaus auch üblich ist, dass Inhalte vor ihrer Veröffentlichung überprüft und manuell freigegeben werden müssen.

Automatisierte Moderationsentscheidungen können sowohl in dem Zeitraum zwischen „Upload“ und Veröffentlichung sowie nachträglich aufgrund einer Meldung erfolgen. Präventive, vor der Veröffentlichung erfolgte Moderationsentscheidungen wurden insbesondere im Bereich des Urheberrechts als „Uploadfilter“ kontrovers diskutiert. Kritiker:innen sehen in ihnen eine „Zensurinfrastruktur“,²² die die Veröffentlichung bestimmter Äußerungen von einer – teilweise rechtlich vorgegebenen – erfolgreichen Prüfung der Rechtmäßigkeit bzw. Vereinbarkeit mit den Nutzungsbedingungen abhängig macht. Automatisierte Überprüfungen können aber auch ältere Inhalte einer Plattform betreffen, etwa nach Anpassungen der Nutzungsbedingungen.

Transparenzberichte und Transparenzdatenbank

Plattformen sind verpflichtet, Moderationsmaßnahmen zu dokumentieren: Online-Plattformen müssen der Kommission die Begründungen der von ihnen durchgeführten Moderationsmaßnahmen melden (Art. 15, 24, 42 DSA). Die Kommission fasst diese Ergebnisse in einer zentralen Datenbank zusammen,²³ welche die Quelle der nachfolgenden Daten ist.

An der Qualität der dort zu erfassten Daten wird Kritik geübt.²⁴ So zeigt sich, dass die verschiedenen Plattformanbieter sehr unterschiedliche Auslegungen der Begrifflichkeiten pflegen. Außerdem melden einige Plattformen wie X keine Maßnahmen. Dies ist bei der

²² Rähm, Jan (2019): Warum Kritiker Angst vor Zensur haben. Deutschlandfunk, <https://www.deutschlandfunk.de/uploadfilter-warum-kritiker-angst-vor-zensur-haben-100.html>.

²³ European Commission (2026): DSA Transparency Database, <https://transparency.dsa.ec.europa.eu>.

²⁴ siehe etwa Groesch, S.; Birrer, A.; Just, N.; Saurwein, F. (2026) Big data, small answers: How the DSA Transparency Database falls short of its regulatory objectives, *Telecommunications Policy* 50(1) <https://doi.org/10.1016/j.telpol.2025.103088>.

Auswertung zu berücksichtigen. Die folgende Auswertung beschränkt sich auf die von den sehr großen Online-Plattformen gemeldeten Begründungen zu ihren Moderationsmaßnahmen und erfasst den Zeitraum 25.09.23 bis 18.02.26. Die Kategorie „Product“ wurde wegen fehlendem Bezug zur Meinungsfreiheit ausgenommen.

Anzahl der Moderationsmaßnahmen

Von den insgesamt rund 5,2 Milliarden Moderationsmaßnahmen gehen fast 44 % (2,3 Milliarden) auf TikTok zurück. Facebook meldet mit etwa 1 Milliarde etwa halb so viel, LinkedIn meldet etwa 827000 Maßnahmen.

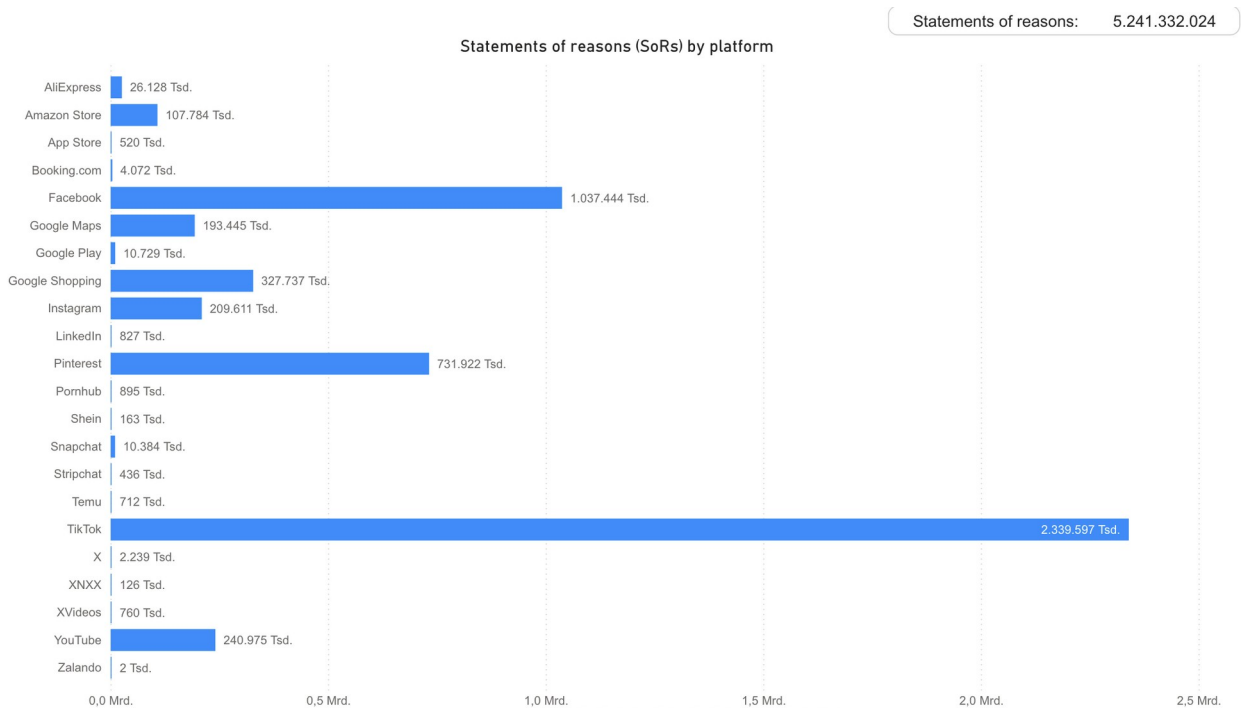


Abb. 1: Statements of Reasons nach Plattform (n=5.241.332.024)(DSA Transparency Database)

Maßnahmentyp

Laut Transparenzdatenbank löschen Plattformen in etwa 46 % der Fälle den Inhalt, während sie in etwa 6,5 % den Zugang sperren, ohne den Inhalt zu löschen. In fast 15 % der Fälle wird der Account suspendiert, in 3,7 % dauerhaft gelöscht. Herabstufungen machen (jedenfalls als explizit zugeordnete Moderationsmaßnahme, Kuratierung ausgenommen) nur etwa 1,5 % der Maßnahmen aus. Knapp 25 % fallen unter die Auffangkategorie „other restriction visibility“, wobei unklar ist, wie großzügig Plattformen diese Kategorie nutzen. Da Herabstufung eine eigene Kategorie darstellt, sollte sie theoretisch nicht hierunter fallen.

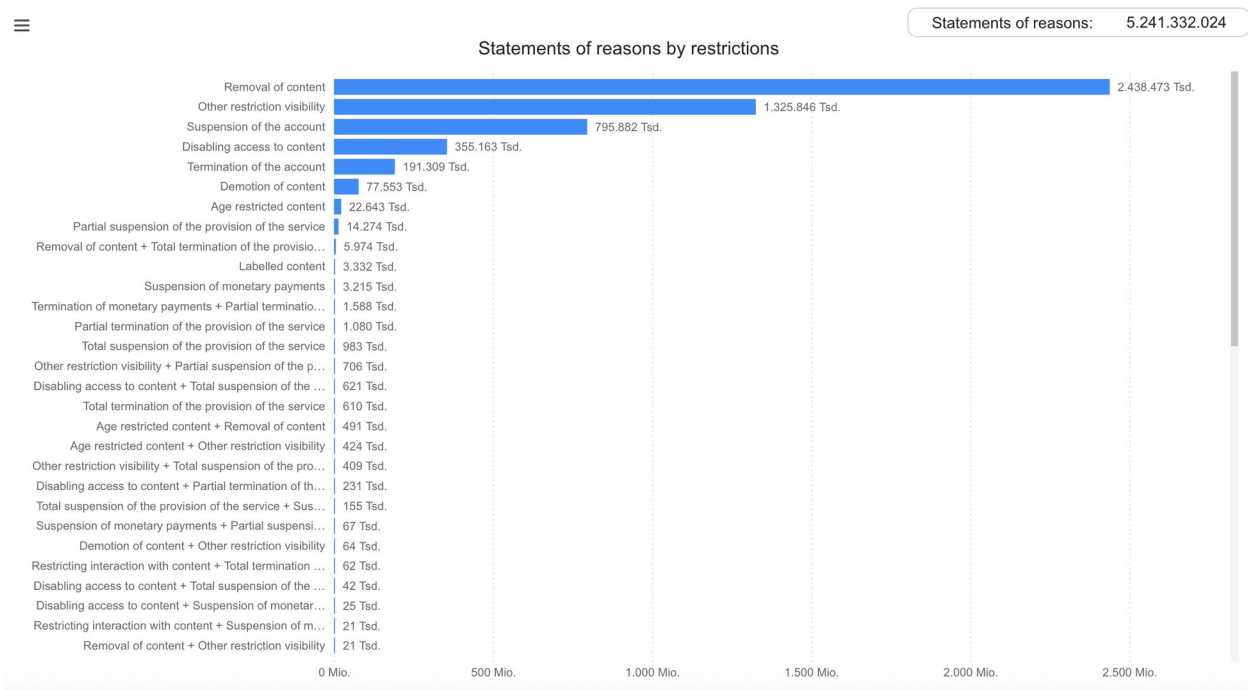


Abb. 2: Statements of Reasons nach Maßnahmentyp (n=5.241.332.024) (DSA Transparency Database)

Erkennung

98,12 % der Moderationsmaßnahmen gehen nicht auf Meldungen durch Dritte zurück, sondern werden durch die Plattformanbieter in Eigeninitiative in Gang gesetzt. Die übrigen rund 1,88 % verteilen sich auf Meldungen durch Nutzende, durch Behörden (Art. 16 DSA) und „vertrauenswürdige Hinweisgeber“ (Art. 22 DSA). Das Meldesystem nach Art. 16 DSA wurde auf VLOPs bislang etwa 18,2 Mio. Mal genutzt (Inhalte der Kategorie „Products“ ausgenommen).

Die Erkennung in Eigeninitiative läuft zu großen Teilen automatisiert. Etwa 3,2 Milliarden, also etwa 61 % aller Maßnahmen, wurden aufgrund automatisierter Erkennung eingeleitet. Dies zeigt den umfassenden Einsatz automatisierter Filtersysteme.

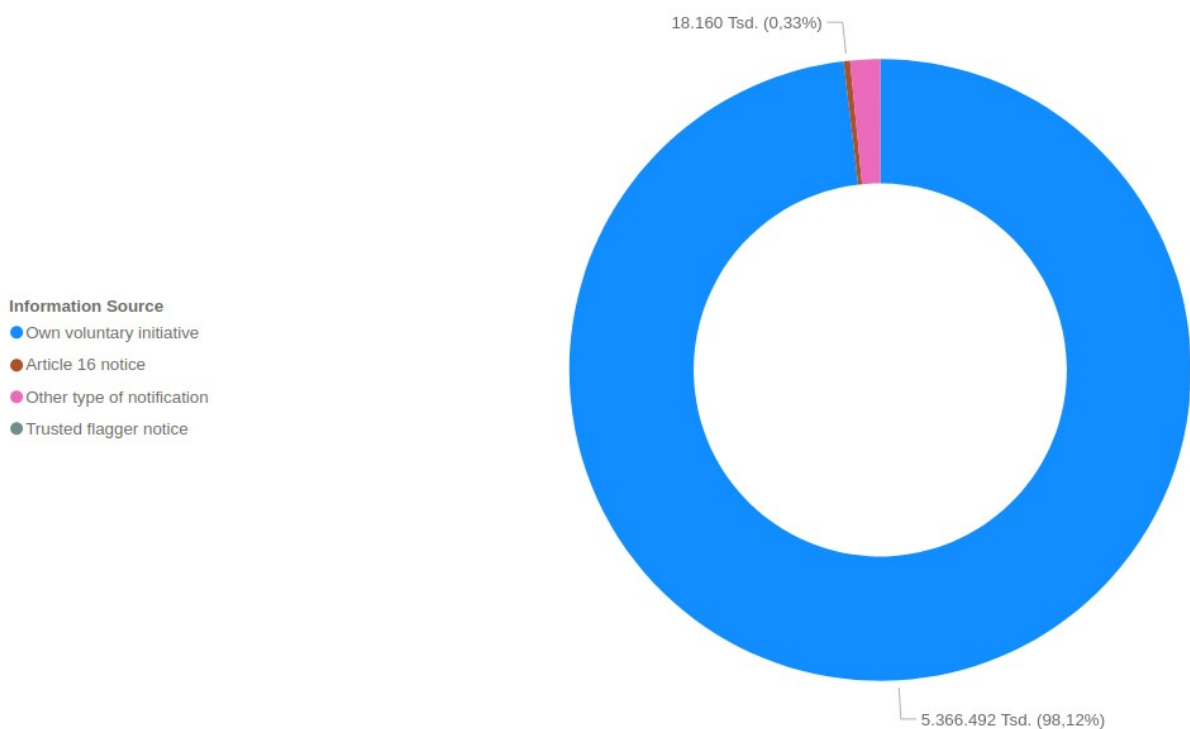
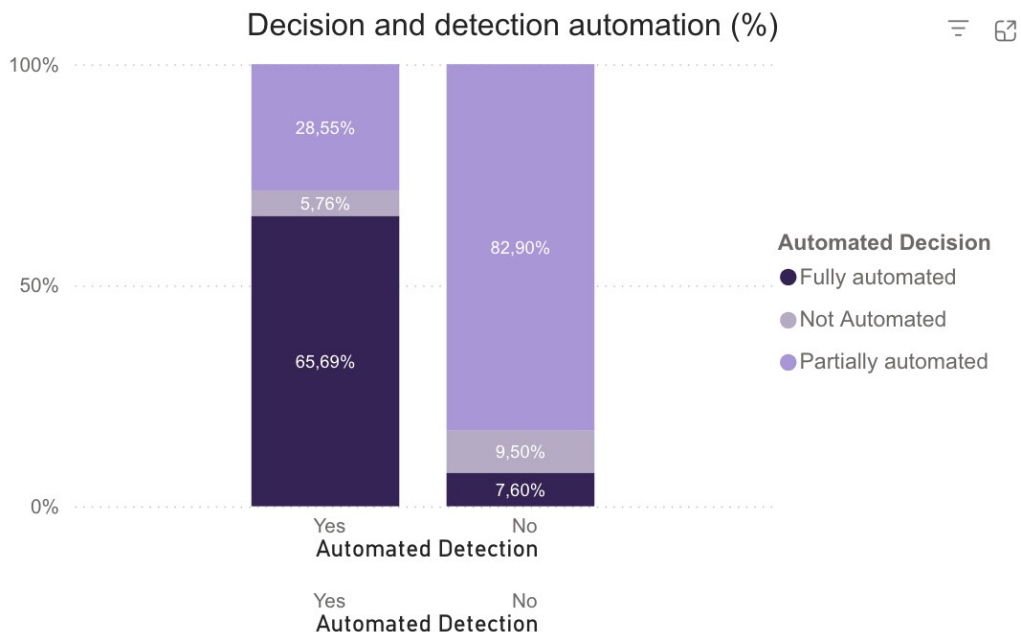


Abb. 3: Informationsquelle (n=5.469.446.852)(DSA Transparency Database)

Entscheidungsmittel

Von der automatisierten Erkennung zu unterscheiden ist, ob auch die Entscheidung automatisiert erfolgte. Dabei unterscheidet die Transparenzdatenbank zwischen den Kategorien voll, teilweise und nicht automatisiert, außerdem nach Art der Erkennung (automatisiert oder nicht automatisiert). Bei automatisierter Erkennung ergingen etwa 66 % der Entscheidungen voll und 29 % der Entscheidungen teilweise automatisiert. Es ist jedoch unklar, was genau unter „teilweise automatisiert“ zu verstehen ist, daher ist es möglich, dass ein erheblicher Teil dieser 29 % dennoch weitgehend ohne menschliche Aufsicht ergeht. Auch bei Verstößen, die nicht automatisiert erfasst wurden, ergingen ca. 8 % der Entscheidungen voll und 83 % der Entscheidungen teilautomatisiert. Nur in 6 % (automatisierte Erkennung) beziehungsweise 10 % der Fälle ergingen die Entscheidungen nicht automatisiert. Auch im Falle einer durch Moderator:innen getroffenen Entscheidung muss beachtet werden, dass diese unter hohem Zeitdruck arbeiten und daher kaum zu ernsthaften Abwägungen in der Lage sind.²⁵



²⁵ Farah, Hibaq (2023): Diary of a TikTok moderator: 'We are the people who sweep up the mess', The Guardian abgerufen am 16.02.2026, <https://www.theguardian.com/technology/2023/dec/21/diary-of-a-tiktok-moderator-we-are-the-people-who-sweep-up-the-mess>.

Abb. 4: Automatisierungsgrad der Entscheidungen nach automatisierter Erkennung (n=5.241.332.024) (DSA Transparency Database)

Entscheidungsgründe

99,53 % der Maßnahmen ergehen wegen Verstößen gegen die AGB, während nur 0,17 % wegen Rechtswidrigkeit erfolgen. Die Transparenzdatenbank enthält auch Angaben zu den Entscheidungsgründen, die sich auch für die einzelnen Plattformanbieter darstellen lassen. Demnach zeigt sich in den plattformübergreifenden Daten, dass etwa 40 % der Maßnahmen mit Verstoß gegen den „scope of platform service“ begründet werden – also einer Unvereinbarkeit mit der Ausrichtung des Dienstes. Hier zeigt sich, wie durch die Verwendung vager Kategorien verdeckt wird, worauf genau sich Moderationsmaßnahmen stützen – den Begründungsanforderungen des Art. 17 DSA genügt eine derart allgemeine Entfernung nicht. Weiter werden etwa 19 % mit der Begründung „illegal or harmful speech“ entfernt. Es folgen „scams/fraud“ (ca. 9,9 %) und other violations of terms and conditions (ca. 9,6 %). Bei etwa 8,3 % lautet die Begründung „pornographic or sexualized content“. Die an die Risikokategorien des DSA erinnernde Kategorie „Negative effects on civic discourse and elections“ rangiert bei 1,5 %. Während prozentual gering, stellt dies fast 80 Mio. Maßnahmen dar, ist daher immer noch von wesentlicher Bedeutung.

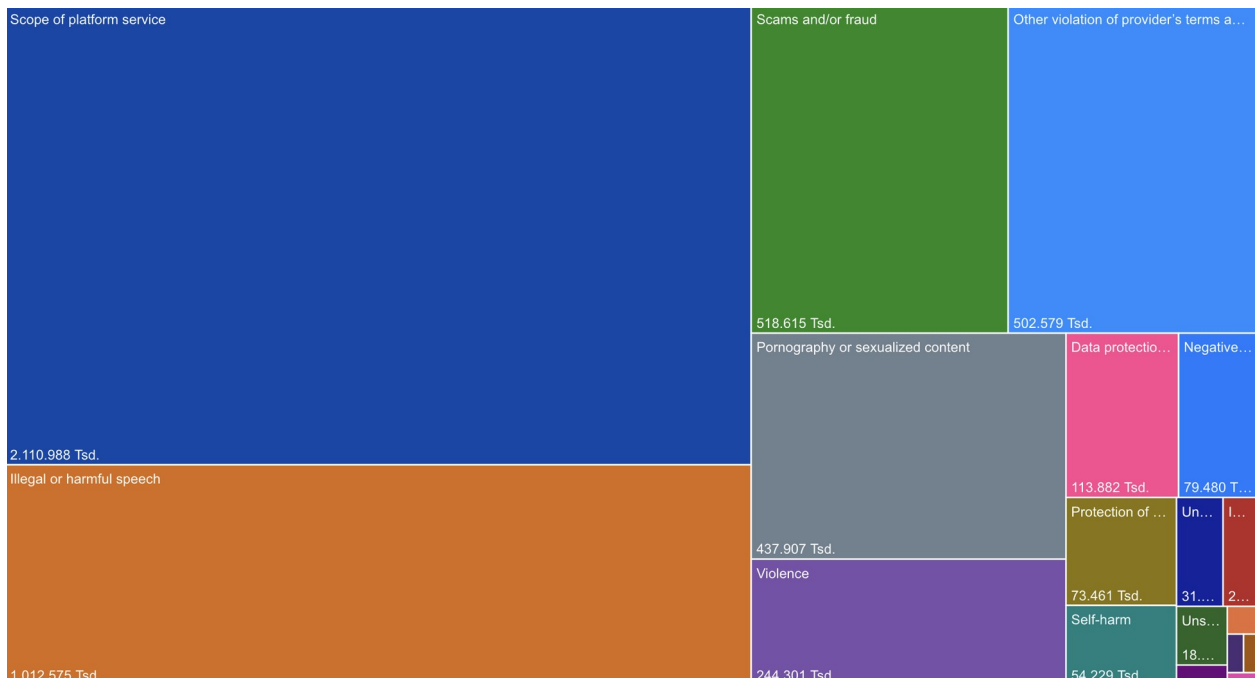


Abb. 5: Begründung der Entscheidung (n=5.241.332.024)(DSA Transparency Database)

4. Wie bewerten Sie das Vorgehen der Anbieter der Online-Plattformen gegen Desinformationen?

Bei der Bewertung des Vorgehens von Online-Plattformen ist zunächst zwischen Desinformation, die rechtswidrig ist, sowie solchen Inhalten, die nicht gegen staatliches Recht verstoßen, zu unterscheiden. An Moderationsmaßnahmen, die sich gegen nicht-rechtswidrige, folglich legale und von der Meinungsfreiheit geschützte, Inhalte richten, sind substantiell strengere Maßstäbe anzulegen. Hierbei ist ferner zu differenzieren zwischen Tatsachen- und Meinungsäußerungen sowie der Gefährlichkeit des Inhalts für Dritte, insbesondere im Hinblick auf Leib und Leben.

Rechtswidrige Desinformation und Verpflichtungen der Plattformen

Rechtswidrige Desinformation, also solche Inhalte, die gegen geltendes deutsches Recht verstoßen, hat regelmäßig einen Personenbezug. Dabei handelt es sich primär um strafbare Verleumdungen oder üble Nachrede (§§ 186 – 188 StGB). Im Bereich der Strafbarkeit der Verbreitung objektiver Falschinformationen ist die Leugnung der nationalsozialistischen Gewalt- und Willkürherrschaft (§ 130 Abs. 3 StGB) besonders hervorzuheben. Über strafrechtlich relevante Inhalte hinaus kann es sich jedoch auch um (zivilrechtliche) Verletzungen absoluter Rechte Dritter handeln, insbesondere von Persönlichkeits-, Marken- oder Namenrechten. Zu denken ist hierbei beispielsweise an nicht-verleumderische Falschinformationen über eine Person und Fälschungen seriöser Nutzerkonten sowie Webseiten (beispielsweise im Rahmen der sogenannten „Doppelgänger“-Kampagne).²⁶

Bezüglich all dieser illegalen Inhalte bestehen jedenfalls Verpflichtungen der Plattformen, die Inhalte nach Meldung zu entfernen, in Teilen sogar ihre Veröffentlichung präventiv zu verhindern.²⁷ Ein Vorgehen der Online-Plattformen gegen solche rechtswidrige

²⁶ Rohwedder, Wulf (2024): Doppelgänger - gekommen und geblieben, Tagesschau, <https://www.tagesschau.de/faktenfinder/kontext/russland-desinformation-analyse-102.html>.

²⁷ EuGH (2019): Glawischnig-Piesczek, Fall C-18/18; OLG Frankfurt a.M. (2024): Künast, Az. 16 U 65/22; für die nun auch angenommene datenschutzrechtliche Verpflichtung zur Verhinderung der Veröffentlichung rechtswidriger Informationen, EuGH (2025): Russmedia, Fall C-492/23; Tuchtfeld, Erik (2025): Eine allgemeine Verpflichtung zur Überwachung: Wie der EuGH das Haftungsregime der europäischen Plattformregulierung auf den Kopf stellt, Verfassungsblog, <https://verfassungsblog.de/eugh-russmedia-digital-uberwachung>.

Desinformation ist zu begrüßen. Die Legitimationsgrundlage für die Maßnahmen der Online-Plattformen liegt hierbei in einem demokratisch beschlossenen Parlamentsgesetz, welches der verfassungsgerichtlichen Kontrolle unterliegt. Das Handeln der Online-Plattformen ist somit – zumindest mittelbar, über seinen normativen Anknüpfungspunkt – in eine gesellschaftliche Debatte zur Zweckmäßigkeit des Verbots der Äußerung im Rahmen des Erlasses des Gesetzes sowie der rechtsstaatlichen Kontrolle bezüglich der Verhältnismäßigkeit des Eingriffs in die Meinungsfreiheit eingebettet.

Legale Desinformation und Maßnahmen der Plattformen

Grundsätzlich anders verhält es sich bei Maßnahmen gegen Desinformation, die nicht gegen staatliches Recht verstoßen. Hierbei fehlt die einer rechtsstaatlichen Kontrolle unterliegende demokratische Kontrolle des Verbots bestimmter Äußerungen. Nichtsdestotrotz kann ein Vorgehen der Online-Plattformen gegen solche Inhalte wünschenswert sein, sofern sie eine erhebliche Gefahr für die Rechte Dritter darstellen. Dabei gilt, dass eine Äußerung desto weniger schützenswert ist, je größer ihr (nachweislich falscher) Tatsachenkern sowie ihre Gefährlichkeit für Dritte ist.

Nach der ständigen Rechtsprechung des Bundesverfassungsgerichts sind Tatsachenbehauptungen vom Schutzbereich der Meinungsfreiheit umfasst, wenn sie Voraussetzung für eine freie Meinungsbildung sind.²⁸ Nicht erfasst sind dagegen bewusst unwahre Tatsachenbehauptungen, weil eine unrichtige Information nicht zum Meinungsbildungsprozess beitragen kann.²⁹ Im Bereich der Desinformation ist daher ein gradueller Schutz von Äußerungen vorzunehmen, bei der die Bedeutung und der Umfang des tatsächlichen Bezugspunkts zu identifizieren ist.

Verhältnismäßige Maßnahmen und Missbrauchsrisiken

Äußerungen, die nahezu ausschließlich Elemente der Stellungnahme und des Dafürhaltens bezüglich einer bestimmten Position enthalten – und die nicht aus anderen Gründen rechtswidrig sind –, sollten regelmäßig nicht von Online-Plattformen moderiert werden, weil sie grundsätzlich zu einer freien Debatte zu einem bestimmten Thema beitragen.

²⁸ BVerfG (2012), 1 BvR 901/11, Rn. 18–20 (mit weiteren Nachweisen).

²⁹ BVerfG (2011), 1 BvR 917/09, Rn. 18 (mit weiteren Nachweisen).

Dies gilt umso mehr bei Äußerungen, die im besonderen Maße auf den öffentlichen Diskurs ausgerichtet sind, beispielsweise aufgrund ihres sozial- oder machtkritischen Inhalts und der Auseinandersetzung mit dem Handeln staatlicher Institutionen. Je größer jedoch der (nachweislich falsche) Tatsachenanteil einer Äußerung wird, desto eher entfällt ihre Schutzwürdigkeit beziehungsweise stellt eine Moderationsmaßnahme gegen die Äußerung einen gerechtfertigten Eingriff in die Meinungsfreiheit dar.

In der Praxis stellen die Anforderungen an die Nachweisbarkeit der Falschheit einer Information eine besondere Herausforderung dar. Hierbei ist insbesondere auf Positionen und Einordnungen staatsferner, regierungsunabhängiger Institutionen abzustellen, beispielsweise der Presse oder wissenschaftlicher Einrichtungen. Widerspricht eine Äußerung einem breiten wissenschaftlichen Konsens zu einem Thema, so darf regelmäßig davon ausgegangen werden, dass sie nachweislich falsch ist. Insoweit Äußerungen aber lediglich von einer wissenschaftlichen Mehrheitsmeinung abweichen, sich aber mit einer relevanten Mindermeinung decken, so sind Moderationsmaßnahmen deutlich kritischer zu beleuchten.

Ferner ist die Gefährlichkeit einer Äußerung für Dritte zu berücksichtigen. So können Lügen über Personengruppen zu Hass aufstacheln und zu schwersten Gewaltverbrechen beitragen. An dieser Stelle sei insbesondere an den Völkermord an den Rohingya in Myanmar erinnert, der auch auf das vollständige Versagen von Sicherheitsmechanismen in der Moderation von Inhalten bei Facebook, betrieben von Meta, zurückzuführen ist.³⁰ Darüber hinaus kann eine besonders hohe Gefährlichkeit für Dritte auch durch falsche Gesundheitsinformationen entstehen. Zu denken sei hierbei beispielsweise an den Vorschlag des US-Präsidenten Trump, Bleiche gegen das Corona-Virus zu trinken.³¹ Auch weniger prominente „Gesundheits-Influencer“ fallen immer wieder mit zweifelhaften, gefährlichen Vorschlägen auf.³²

³⁰ Amnesty International (2022): Myanmar: Facebook-Algorithmen haben Gewalt gegen Rohingya befördert, <https://www.amnesty.de/allgemein/pressemitteilung/myanmar-facebook-algorithmen-haben-gewalt-gegen-rohingya-befoerdert>.

³¹ Tagesschau (2020): Desinfektionsmittel? Alles nur „Sarkasmus“, <https://www.tagesschau.de/ausland/trump-desinfektionsmittel-101.html>.

³² WDR (2025): TikTok-Heiler: Eine kaum kontrollierte Gefahr, <https://www1.wdr.de/nachrichten/landespolitik/medfluencer-gefaehrliche-gesundheitstipps-100.html>;

Als weitere Risiko-Kategorie kommt auf Online-Plattformen deren Verbreitungslogik und damit die potenzielle Reichweite von Desinformation hinzu. Am analogen Stammtisch ist das Schadenspotenzial einer bewussten Falschinformation gering, auf Online-Plattformen erreicht eine entsprechend aufbereitete Falschinformation jedoch potenziell ein Millionenpublikum. Der nötige Aufwand an Sorgfalt der Plattformen bei der Prüfung eines sich „viral“ verbreitenden Inhalts erscheint vor diesem Hintergrund deutlich höher, als ein gewisses „Grundrauschen“ potenziell desinformierender Inhalte.

Mehr Zurückhaltung ist unserer Meinung nach dagegen bei Äußerungen erforderlich, die nicht unmittelbar gefährlich sind, sondern lediglich indirekt nachteilige Effekte für die Gesundheit Dritter entfalten können, beispielsweise indem sie das Vertrauen in relevante Institutionen (Gesundheitsämter, Forschungseinrichtungen, medizinisches Personal etc.) unterminieren. Insbesondere bei staatlichen und supra-staatlichen Einrichtungen sind die Äußerungen regelmäßig, mögen sie in der Sache auch falsch oder verzerrend sein, als zulässige Machtkritik zu verstehen, die in einer liberalen Demokratie möglich sein muss. Einrichtungen dieser Art haben in der Regel besondere Möglichkeiten auf solche Vorwürfe zu reagieren, beispielsweise über eigene Kommunikationsmittel oder durch Darstellung ihrer Perspektive in Qualitätsmedien. Auch scharfe und unsachliche Kritik ist daher regelmäßig auszuhalten.

Angemessen können jedoch Moderationsmaßnahmen unterhalb der Löschung des Inhalts sein,³³ sofern diese für den Nutzenden transparent und überprüfbar gemacht werden. Dabei kommt beispielsweise ein algorithmisches Downranking der Inhalte in Betracht, wie es in der Corona-Pandemie erfolgte, die Demonetarisierung von Inhalten oder die Richtigstellung falscher oder verzerrender Inhalte durch Informationsbanner, die unmittelbar neben dem Inhalt angezeigt werden (Faktenchecks). Im Bereich von Faktenchecks zeigen community-

Heckmann, Dirk (2024): Gesundheitsgefährdung durch Social Media?. In: Bergen, I., Gramm, F., Grütters, J., Kolbe, H. (Hrsg.), Wie die Generation Z das Gesundheitswesen verändert, https://doi.org/10.1007/978-3-662-70622-0_7.

³³ Für einen Überblick siehe bspw. Karami, Amir (2025): A dual typology of social media interventions and deterrence mechanisms against misinformation, *Misinformation Review*, <https://misinforeview.hks.harvard.edu/article/a-dual-typology-of-social-media-interventions-and-deterrence-mechanisms-against-misinformation/>.

basierten, die in ihrer Grundlogik oft Parallelen zur Qualitätskontrolle bei Wikipedia haben, und experten-basierten Ansätze unterschiedliche Stärken, stehen jedoch auch nicht im Gegensatz zueinander und können sich ergänzen.³⁴

Auch der Code of Conduct on Disinformation³⁵ zielt als der Teil der Compliance unter dem DSA (Art. 45) primär auf solche „weichen“ Maßnahmen ab, wie die Demonetarisierung von Falschinformationen, die Stärkung von Nutzenden sowie von Wissenschaftler:innen durch Datenzugang.

Abschließend ist zu betonen, dass das Vorgehen gegen Desinformation von hoher Bedeutung für den öffentlichen Debattenraum ist. Nur auf Grundlage eines gemeinsamen Tatsachenkerns lässt sich ein sinnvoller demokratischer Diskurs mit unterschiedlichen politischen Bewertungen führen. Gleichzeitig ist jedoch auch festzustellen, dass es Bestrebungen verschiedener Regierungen gibt, über den Kampf gegen Desinformation oder „Fake News“ den öffentlichen Diskurs einzuschränken und Kritik zu unterbinden. Dies reicht von der Delegitimierung freier und unabhängiger Berichterstattung durch den aktuellen US-Präsidenten bis hin zu Regierungseinheiten, die (vermeintliche) Falschinformationen über Regierungshandeln suchen und in der Folge die Löschung der Inhalte verlangen und die Inhalteersteller strafrechtlich verfolgen lassen.³⁶ Das Vorgehen gegen Desinformation muss deshalb stets mit größter Vorsicht erfolgen und darf insbesondere nicht als Methode der Unterdrückung – auch fernliegender – Meinungen und Kritik missbraucht werden.

³⁴ Für einen Vergleich beider Ansätze Dobusch, Leonhard (2018): Wer checkt die Faktenchecker? Kontroverse um Facebooks „externe Faktenprüfung“, Netzpolitik, <https://netzpolitik.org/2018/wer-checkt-die-faktenchecker-kontroverse-um-facebooks-externe-faktenpruefung/>.

³⁵ European Commission (2025): The Code of Conduct on Disinformation, <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>.

³⁶ Siehe diesbezüglich für Indien Jain, Anmol (2024): The Bombay High Court Dismisses the Ministry of Truth, Verfassungsblog, <https://verfassungsblog.de/india-fact-checking-unit-fake-news>.

5. Tragen die Maßnahmen der Inhalte-Moderation durch Online-Plattformen zur Gewährleistung des Rechts auf freie Meinungsäußerung angemessen bei?

Nein, insgesamt ist derzeit nicht davon auszugehen, dass den Anforderungen der Meinungsfreiheit ausreichend begegnet wird. Auch wenn die Nutzungsbedingungen vieler Plattformen mittlerweile zumindest teilweise hinreichend ausgestaltet sind, mangelt es an einer genügenden Umsetzung dieser Vorgaben sowie der Verpflichtungen des DSA in der Praxis. Die folgenden Problemfelder verdeutlichen dies:

Automatisierte Verfahren

Zunächst fallen Schwächen der automatisierten Verfahren zur Inhaltmoderation auf. Diese sind zumindest derzeit noch nicht ausreichend, um den Nuancen der Meinungsfreiheit hinreichend nachzukommen. Es ist auch nicht davon auszugehen, dass sie dies in absehbarer Zeit sein werden. Außerdem ist es fraglich, ob eine Grundrechtsabwägung in Grenzfällen überhaupt durch algorithmische Systeme vorgenommen werden sollte.

Eine rein menschliche Moderation wäre zwar angesichts der großen Zahl an Inhalten faktisch nicht durchführbar. Besonders problematisch sind jedoch automatische Moderationsverfahren in Beschwerdeverfahren zu beurteilen. Auch wenn Art. 20 Abs. 6 DSA ein rein automatisiertes Beschwerdeverfahren untersagt, gibt es verschiedene Anekdoten aus der Praxis, die darauf hinweisen, dass auch hier offensichtliche Fehlentscheidungen getroffen werden, sodass die menschliche Einbeziehung entweder nicht ausreichend stattfindet oder – bspw. aufgrund Zeitdrucks – ineffektiv ausgestaltet ist. Dafür spricht auch die sehr hohe Quote an Entscheidungen gegen Plattformen vor außergerichtlichen Streitbeilegungsstellen.³⁷

³⁷ European Commission (2026): Two years of Digital Services Act allows 50 million content moderation decisions by platforms to be reversed, <https://digital-strategy.ec.europa.eu/en/news/two-years-digital-services-act-allows-50-million-content-moderation-decisions-platforms-be-reversed>.

Mangelnde Begründungen

Besonders problematisch ist darüber hinaus, dass Moderationsentscheidungen häufig nicht ausreichend begründet sind, obwohl der DSA hierzu sehr konkrete Vorgaben enthält. Das erschwert es Nutzenden, überhaupt zu erkennen, weshalb es zu einer Moderationsmaßnahme kam. Es macht es aber auch unmöglich, Fragen der Reichweite der Meinungsfreiheit bei Plattformen zu diskutieren, sei es in der Öffentlichkeit, in der Forschung oder auch vor Streitschlichtungsstellen nach Art. 21 DSA oder Gerichten.

Obwohl Plattformen die Existenz von Wortlisten leugnen,³⁸ berichten Nutzende von einer gezielten Abwertung ihrer Inhalte durch bestimmte Schlagworte.³⁹ Nutzende haben auf diese Erkennung von Worten reagiert, sodass Kraftausdrücke und bestimmte, politisch potenziell sensible Begriffe verändert oder vermieden werden. Beispiele hierfür sind Verfremdungen wie „Fem!nist“ „F*ck“ oder Begriffe wie „Unalive“ (statt Kill).⁴⁰

Teilweise sind auch Listen bekannt geworden, welche Worte zu von Plattformen als unangemessen angesehen werden.

Relevanz des Empfehlungssysteme

Neben der Inhaltmoderation haben Empfehlungssysteme eine große Relevanz für die Ausübung der Meinungsfreiheit auf Plattformen. So gibt es immer wieder Anzeichen und Berichte, dass diese bestimmte Inhalte besonders stark verteilen, andere hingegen unterdrücken.

Trotz der Transparenzvorgaben des DSA besteht in der Praxis große Intransparenz der tatsächlichen Funktionsweise dieser Systeme, sodass das genaue Ausmaß der Beeinflussung der Meinungsfreiheit schwer zu bestimmen ist. Berichte deuten darauf hin,

³⁸ Steen, Ella, Yurechko, Kathryn, & Klug, Daniel (2023) You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231194586>.

³⁹ Germain, Thomas (2025), The words you can't say on the internet, <https://www.bbc.com/future/article/20251118-the-words-you-cant-say-on-the-internet>.

⁴⁰ Zu sprachlichen Veränderungen aufgrund von Inhalte-Moderation siehe Steen, Ella, Yurechko, Kathryn, & Klug, Daniel (2023) You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231194586>.

dass sich die Moderation von „klassischen“ Maßnahmen hin zu Maßnahmen wie Shadowbanning verschiebt – sei es gezielt oder algorithmisch entwickelt. Unsere Beobachtungen können diese Entwicklung bestätigen.

Erkenntnisse aus außergerichtlicher Streitbeilegung

Kommt es zu einer außergerichtlichen Streitbeilegung, wird dort stark überwiegend den Nutzenden Recht gegeben.⁴¹ Aktuelle Zahlen der Kommission zeigen, dass dies bei außergerichtlichen Streitbeilegungsstellen etwa 52 % der Entscheidungen zu Inhaltsmoderation betraf, bei internen Beschwerdeverfahren etwa 30 %.⁴² Das bedeutet zwar nicht, dass derart viele Moderationsentscheidungen generell falsch sind. Zugleich ist es alarmierend, da in diesen Fällen zunächst über eine interne Beschwerde entschieden wurde und auch im Rahmen der außergerichtlichen Streitbeilegung die Möglichkeit besteht, eine erneute Entscheidung zu treffen.

Problematisch ist hierbei auch, dass diese Entscheidungen die Plattformen nicht binden. Derzeit wird offenbar nur etwa die Hälfte der Entscheidungen tatsächlich umgesetzt. Soziale Plattformen sind heute zentrale Infrastrukturen der öffentlichen Meinungsbildung, das macht ihre Intransparenz zu einem Problem für die Demokratie. Dabei ist vor allem das Zusammenwirken verschiedener Mechanismen problematisch: Algorithmen entscheiden im Verborgenen, welche Inhalte Reichweite erhalten und welche unsichtbar bleiben, Moderationsmaßnahmen wie Shadowbanning bleiben von Nutzenden unbemerkt und damit kaum anfechtbar und unabhängige Forschung wird durch fehlenden Datenzugang verhindert. Indem sie bestehenden rechtlichen Verpflichtungen einseitig oder nur vereinzelt nachkommen, entziehen sich große Plattformen systematisch der Kontrolle. Daher wird dem Recht auf freie Meinungsäußerung im digitalen Raum noch nicht genug Rechnung getragen.

⁴¹ Appeals Centre Europe Transparency Report (2025): November 2024 to August 2025, <https://www.appealscentre.eu/wp-content/uploads/2025/09/Appeals-Centre-Europe-Transparency-Report.pdf>

⁴² European Commission (2026): Two years of Digital Services Act allows 50 million content moderation decisions by platforms to be reversed, <https://digital-strategy.ec.europa.eu/en/news/two-years-digital-services-act-allows-50-million-content-moderation-decisions-platforms-be-reversed>.